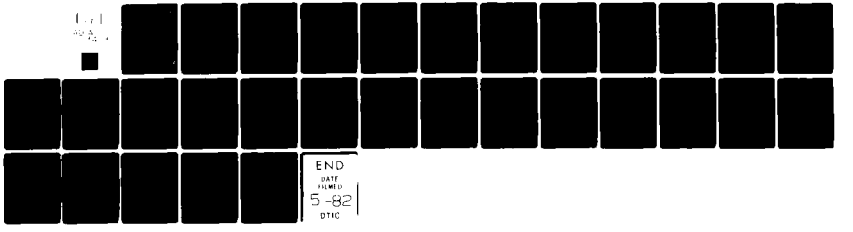
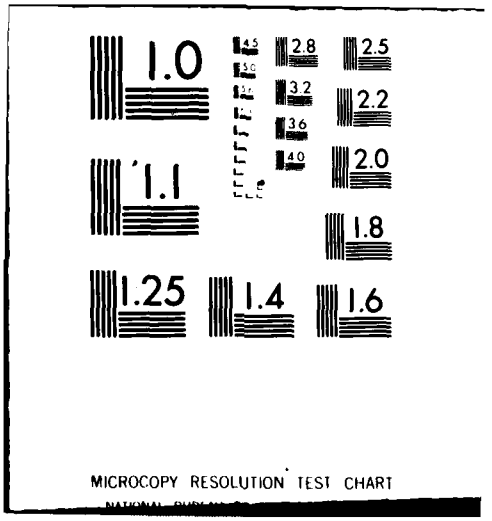
AD-A113 419 STANFORD UNIV CA DEPT OF COMPUTER SCIENCE F/G 12/1
FAST MATRIX MULTIPLICATION WITHOUT APA-ALGORITHMS,(U)
OCT 81 V PAN N00014-81-K-0269
UNCLASSIFIED STAN-CS-81-882 NL



END
DATE FILMED
5-82
DTIC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
MICROCOPY RESOLUTION TEST CHART

October 1981

Report. No. STAN-CS-81-882

5

AD A113419

Fast Matrix Multiplication Without APA-Algorithms

by

V. Pan

Department of Computer Science

Stanford University
Stanford, CA 94305

Contract N00014-76-C-0322

DTIC FILE COPY

DTIC
ELECTE
APR 14 1982
S D
D

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

Fast Matrix Multiplication Without APA-Algorithms

V. Pan
Computer Science Department
Stanford University
Stanford, California 94305

Abstract. The method of trilinear aggregating with implicit canceling for the design of fast matrix multiplication (MM) algorithms is revised and is formally presented with the use of Generating Tables and of linear transformations of the problem of MM. It is shown how to derive the exponent of MM below 2.67 even without the use of approximation algorithms.

Key Words: bilinear algorithms, direct sum, matrix multiplication, tensor product construction, trilinear aggregating.

Accession For	
NTIS GRA&I	☒
DTIC TAB	☐
Unannounced	☐
Justification	
By Per Ltr. on file	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A	

This research has been supported by National Science Foundation grant MCS-77-23738 and Office of Naval Research contract N00014-81-K-0269 at Stanford University, by National Science Foundation grant MCS-8003347 at SUNY, Albany. Reproduction in whole or in part is permitted for any purpose of the United States government.

1. Introduction.

The attention to the problem of fast matrix multiplication hereafter referred to as MM has remained very high since 1968 when V. Strassen proved that $4.8N^{2.81}$ arithmetic operations rather than $2N^3$ suffice to multiply two $N \times N$ matrices for all N, see [1]. (For comparison, a method of nonasymptotic acceleration of MM [2] presented in January 1966 at the seminar of Dr. G. M. Adelśon-Velśkii, Dr. A. S. Kronrod, and Dr. Y. M. Landis in Moscow has not been published because of the lack of interest to that method outside the seminar in 1966.)

The attempts to improve the exponent 2.81 followed. Smaller exponents could automatically result from any sufficiently fast (in terms of the number of nonscalar multiplications involved) bilinear algorithm for a MM problem of any specific shape because of the possibility to use bilinear algorithms recursively. (Hereafter that number of nonscalar multiplications is called the multiplicative complexity of a bilinear algorithm.) The design of fast basic algorithms for the recursion turned out to be a harder problem. The next improvement of the exponent from 2.81 to 2.7804 came about in 1978, see [3]. The proof techniques (trilinear aggregating, uniting and canceling, TAUC) have been sketched in the earlier paper [4]. However the actual potential power of the TAUC has not been fully appreciated even in 1978. Later another approach to the acceleration of MM (called the method of APA-algorithms) appeared in [5] and has been justified in [6]. This reduced the exponent to 2.7799. Then the methods of APA-algorithms and TAUC have been combined together which led to a more serious asymptotic acceleration of MM, see [7-10]. On the other hand, it turned out that the TAUC are closely related to the Direct Sum Problem (DSP) of the fast evaluation of (the direct sum of) r independent sets of bilinear forms, $r > 1$. According to the Direct Sum Conjecture (DSC), due to [11], the multiplicative complexity of the whole problem (Direct Sum Problem) is equal to the sum of multiplicative complexities of the r independent problems of the evaluations of r given sets. On the contrary, the TAUC successfully exploits the advantage of simultaneous evaluation of several independent sets of matrix products. In the case of APA-algorithms the TAUC enables us to disprove the DSC. The first formal counterexample to the DSC over the class of APA-algorithms appeared in [8, Remark, p. 37] although the DSC has not been studied in [8]. (See other counterexamples also based on the TAUC in [9,10].) In the case of usual algorithms the DSP remains open. This might be due to our poor knowledge of the lower bounds. For example, no method is known for 10×10 MM in 650 nonscalar multiplications while two 10×10 matrix products can be evaluated using the TAUC in 1300 multiplications. (However this does not disprove the DSC because the best known lower bound for 10×10 MM is only 199 multiplications.) Of course, the latter algorithm for the pair of 10×10 MM can be transformed into a fast algorithm for 10×10 by 10×20 MM in 1300 multiplications which can be applied as a basis for the recursion to derive an asymptotically fast method for MM. On the other hand, the recursion based on the method of 10×10 MM in 650 would result in a smaller exponent of MM. Although, as we mentioned, such a method for 10×10 MM might not exist it turned out that practically the same exponent can be obtained as if it existed because the recursion can also be used with an algorithm for a direct sum of MM problems as a basis. A similar result for any basic algorithm for an arbitrary direct sum of MM problems is due to [10] and is known as the Exponential Direct Sum Theorem, EDST; see [9]. It is worth mentioning that historically the earlier techniques of the TAUC motivated the EDST as a means to reinforce the power of the

TAUC.

By combining the new methods of the TAUC and APA-algorithms with each other, with the EDST, and with the recursive construction (which is also called the Tensor Product Construction (TPC)) smaller exponents of MM were obtained in 1979; see [7–10]. (The references to the TAUC are omitted in [7, 10] but the reader can easily notice common basic elements of the patterns of [7, 10] and of the earlier 2-Procedure of the TAUC of [3, 4, 12]; see also [8], [9, Section 19], and [13, Section 4].) In particular, the exponent 2.522 was obtained by combining the construction of [8] with the EDST and was announced on October 26, 1979, at the Conference on the Complexity Theory in Oberwolfach, October 21–27, 1979 (see [14]) although only out-of-date 2.548 appeared as the "world record" in the EATCS report on that conference [15]. Later improvements in 1980–81, see [9, 10, 16, 17], which reduced the exponent to 2.5167, 2.5161, 2.496 also relied on the combinations of the techniques of the TAUC, APA-algorithms, EDST, TPC, and on some new elements of the analysis. However in general the progress seems to go out of power after 1979 because the most natural combinations of that kind have already been explored. (So called Partial Matrix Multiplication technique, see [7], does not seem to lead to a serious if any improvement over the EDST.)

We believe that the further progress in the acceleration of MM and might be in the solution of the DSP for usual algorithms depends on the success in the analysis of the methods of trilinear aggregating (TA) because TA constitutes the basis for the design of the fastest MM algorithms. This paper is our extensive attempt of such an analysis. Thus we intentionally focus our attention on TA.

We formally define the process of TA by reducing it to the design of Generating Tables which in turn are obtained from certain partitions of finite sets. Until the last section we do not involve APA-algorithms because we tend to simplify the problem and to understand how successfully TA can work without them. Our study shows that the resulting MM algorithms are quite fast even if APA-algorithms are not used. On the other hand, the structure of our algorithms is more regular than the structure of the faster algorithms for MM obtained via the APA-algorithms.

To make the paper self-contained we formally state the problem of MM and of the direct sum of MM and prove the EDST in Sections 2 and 3. In our proof we follow [9] using Theorem 13.1 of [9] as a basis but the successful notation borrowed from [10] helped us to make the proof much simpler. (Formally we prove the EDST for usual algorithms. The extension to the case of APA-algorithms is well understood now; see [6, 9, 10, 17].) Our proof of the EDST unlike the proofs of [10, 17] is elementary and does not use tensorial calculus. Also in Section 2 we show that the asymptotic complexity of MM can not depend on the choice of the field of constants unless such a field is finite. In Section 4 we revisit the TAUC. We present it more formally than we did earlier and in a different version. The procedures of trilinear aggregating (TA) and consequently MM algorithms are defined by Generating Tables (GT). The resulting algorithms for MM appear as decompositions of special trilinear forms (associated with the given problems of MM) into sums of aggregates and correction terms obtained from the Generating Tables. The total number of terms equals the multiplicative complexity of the algorithms and consequently defines the exponents of MM. Hence our objective is the reduction of the total number of terms and, in particular, of the number of the correction terms because the aggregates are not numerous.

In Section 5 we rewrite the GTs so that the design of algorithms for larger problems of MM appears in a more explicit fashion than in the cases where it is defined by the recursive process that starts with the algorithms for small MM problems. Also we define the degree and the dimension of correction terms of a Generating Table and show why it is desirable that all of or most of the correction terms have degree 1. In Section 6 we show that the latter property follows if the GTs are defined by some appropriate partitions of the finite sets. We give two examples of the GTs (the First and the Second Constructions of Section 6) where we demonstrate which properties of the partitions are to be exploited. In Section 7 we describe the method of Implicit Canceling (IC) of correction terms of degree 1; see [13], to be combined with TA to define Trilinear Aggregating with Implicit Canceling (TAIC). TAIC is a modification of TAUC. It provides us with an insight into the techniques of the design of fast MM algorithms. Combining TAIC with the First Construction of Section 6 gives us a quite regular and homogeneous algorithm that evaluates (the direct sum of) $(2u)!/(u!)^2$ independent products of $n^u \times n^{2u}$ by $n^{2u} \times n^u$ matrices in $(n+1)^{4u}$ multiplicative steps for arbitrary natural n and u. This defines the exponents less than 2.67 without the use of auxiliary APA-algorithms. (The best previous result of that kind was 2.773 ...; see [13].) Combining TAIC with the Second Construction of Section 6 gives a similar method for the direct sum of $(3v)!/(v!)^3$ independent problems of $(n-1)^{3v} \times (n-1)^{3v}$ MM involving $(n+1)^{9v}$ multiplicative steps for arbitrary natural n and v. This defines the exponents less than 2.7288 (also without the use of APA-algorithms.) Technically the latter algorithm involves TAUC and a method of Alternating Summation of Aggregates which is used to cancel the terms of positive codimensions. Finally in Section 8 we sketch the possible generalizations of our approach. This includes the study of the partitions of finite sets for GTs (with the First and Second Constructions of Section 6 as the models) and of the Generating λ-Tables. In the latter case the indeterminates appear in the GTs with some constant coefficients which may depend on a parameter λ. This case incorporates TAUC with a special Canceling Procedure (see [3, 12]) and the design of APA-algorithms which are sometimes also called λ-algorithms (see [8, 9, 17]).

We hope that our analysis will help the reader to understand the principles of trilinear aggregating (which we consider the basic technique for fast MM) and finally will lead to a new acceleration of MM in the future.

2. Some Basic Notions, Basic Notation, Basic Construction.

Hereafter $u_{ik} = (U)_{ik}$ designates the (i,k) entry of a matrix U, $\underline{U}$ designates a vector of all entries of U taken in a fixed order, $\mathrm{Tr}\, U = \sum_\ell u_{\ell\ell}$ is the trace of U. I, J, K are given natural numbers, i, j, k are integer parameters.

Definition 2.1. $\langle I,J,K\rangle$ the problem of MM. Given a field (of constants) F, an $I \times J$ matrix X, and a $J \times K$ matrix Y whose entries are indeterminates. Evaluate (the entries of) the product XY by a straight line arithmetic algorithm using the constants from F.

$\langle I,J,K\rangle$ is an example of a bilinear arithmetic computational problem that is the problem of the evaluation of a given set of bilinear forms, $\mathcal{B}$. In the case of $\langle I,J,K\rangle$, $\mathcal{B}$ is the set of the entries of XY which are bilinear forms of the entries of X, Y.

In general, a bilinear problem can be equivalently represented by a set of bilinear forms,

$$\mathcal{B} = \{B_\eta(\underline{X}, \underline{Y})\}, \tag{2.1}$$

by a trilinear form

$$T = T(\underline{X}, \underline{Y}, \underline{Z}) = \sum_\eta B_\eta(\underline{X}, \underline{Y}, \underline{Z}) z_\eta, \tag{2.2}$$

or by a tensor $t = (t_{\mu\nu\eta})$ of the coefficients of T; see [4, 18], for surveys on bilinear problems and algorithms, see [19-23].

In the case of $\langle I, J, K \rangle$,

$$T = \mathrm{Tr}(XYZ) = \sum_{i,j,k} x_{ij}\, y_{jk}\, z_{ki}. \tag{2.3}$$

Here X, Y are given matrices to be evaluated (see Definition 2.1) and $Z = (z_{ki})$ is the (auxiliary) $K \times I$ matrix whose entries are indeterminates.

As another example of bilinear problems we mention polynomial multiplication (PM) also known as convolution of vectors (see [21, 23]). PM is defined by the following trilinear form,

$$T = \sum_{i=0}^{p-1} \sum_{j=0}^{q-1} x_i\, y_j\, z_{i+j}. \tag{2.4}$$

Bilinear algorithms for bilinear problems can be equivalently represented as the following bilinear, trilinear or tensorial identical decompositions.

$$\forall \eta: \quad B_\eta(\underline{X}, \underline{Y}) = \sum_{q=1}^{M} f''_{q\eta}\, L_q(\underline{X})\, L'_q(\underline{Y}), \tag{2.5}$$

$$T(\underline{X}, \underline{Y}, \underline{Z}) = \sum_{q=1}^{M} L_q(\underline{X})\, L'_q(\underline{Y})\, L''_q(\underline{Z}), \tag{2.6}$$

$$t_{\mu\nu\eta} = \sum_{q=1}^{M} f_{q\mu} f'_{q\nu} f''_{q\eta} \quad \text{for all } \mu, \nu, \eta. \tag{2.7}$$

Here

$$\forall q: \quad L_q(\underline{X}) = \sum_\mu f_{q\mu} x_\mu,\ L'_q(\underline{Y}) = \sum_\nu f'_{q\nu} y_\nu,\ \ L''_q(\underline{Z}) = \sum_\eta f''_{q\eta} z_\eta, \tag{2.8}$$

$$f_{q\mu}, f'_{q\nu}, f''_{q\eta} \in F \quad \text{for all } q, \mu, \nu, \eta. \tag{2.9}$$

Hereafter the reader may identify a bilinear algorithm with either of its three representations but actually the evaluation of $\mathcal{B}$ proceeds by first computing the M products $\pi_q(\underline{X}, \underline{Y}) = L_q(\underline{X}) L'_q(\underline{Y})$ for all q, and then computing $B_\eta(\underline{X}, \underline{Y}) = \sum_{q=1}^{M} f''_{q\eta} \pi_q(\underline{X}, \underline{Y})$ for all η. Hereafter M is called *the rank of a bilinear algorithm.*

In the case of MM the subscripts μ, ν, and η are represented by the pairs of (i, j), (j, k), and (k, i) respectively (for example, in such a case $y_\nu = y_{jk}$, $f''_{q\eta} = f''_{qki}$).

We will refer to the tensorial representation (2.7) in Remark 2.1 but otherwise the reader may skip (2.7). In fact, we presented the tensorial representation only for the sake of completeness because of its wide use in the literature on MM. Furthermore we will need only the trilinear representation after Section 3.

The equivalence of (2.5), (2.6) and (2.7) is easily verified. For instance, for the transition from (2.6) to (2.5) equate the coefficients of each indeterminate z_η in the left and right sides of (2.6). Equating the coefficients of all x_μ or of all y_ν rather than z_η we obtain the two (dual) bilinear algorithms of the same rank M for the two dual bilinear problems $\{B_\mu(\underline{Y}, \underline{Z})\}$ and $\{B_\nu(\underline{Z}, \underline{X})\}$.

For example, if the original algorithm of rank M solves $\langle I, J, K\rangle$ then the dual ones solve $\langle J, K, I\rangle$ and $\langle K, J, I\rangle$ and have the same rank, M. In fact, such algorithms can be also transformed into ones of the same rank, M, for the problems $\langle J, I, K\rangle$, $\langle I, K, J\rangle$, and $\langle K, J, I\rangle$. (Indeed, substitute u_{ji}, v_{ik}, w_{kj} for x_{ij}, z_{ki}, and y_{jk} respectively in (2.3) and (2.6).) The study of the asymptotical time-complexity of bilinear algorithms for MM relies on the next definition and theorem.

Definition 2.2. $\beta = \beta(F)$ *is an exponent of* MM (over F) if there exists a positive constant $c = c(\beta)$ associated with that exponent β such that cN^β arithmetic operations are sufficient to solve $\langle N, N, N\rangle$ for all N by straight line algorithms (with the constants from F). β^* is a *limiting exponent of* MM if for all $\epsilon > 0$, $\beta^* + \epsilon$ is an exponent of MM.

Theorem 2.1; see [1]. *If for some natural numbers I, J, K, M there exists a bilinear algorithm (2.5) (2.9) of rank M for $\langle I, J, K\rangle$ then $\beta = 3\log M/\log(IJK)$ is an exponent of MM.*

Outline of Proof. The *basic observation* for the proof is that in the case of MM the identities (2.5)–(2.9) remain true if the entries of $\underline{X}$, $\underline{Y}$, $\underline{Z}$ are replaced by the $I' \times J'$, $J' \times K'$ and $K' \times I'$ matrices respectively (for arbitrary I', J', K'). Then $L_q(\underline{X})$, $L'_q(\underline{Y})$, $L''_q(\underline{Z})$ for all q are also $I' \times J'$, $J' \times K'$ and $K' \times I'$ matrices respectively and $\mathrm{Tr}(XYZ)$ represents $\langle II', JJ', KK'\rangle$. If $I = J = K$ we write $I' = J' = K' = I$ and apply the original algorithm to multiply $L_q(\underline{X})$ by $L'_q(\underline{Y})$ for all q. This defines the transition from a bilinear algorithm of rank M for $\langle I, I, I\rangle$ to the one of rank M^2 for $\langle I^2, I^2, I^2\rangle$. Continuing this process and counting the number of arithmetic operations we obtain the desired upper bound in the cases $N = I^h$ for all h and then easily extend the bound to the case of arbitrary N. If (I, J, K) is an arbitrary triplet we come back to the square MM by writing $I' = J$, $J' = K$, $K' = I$ and then $I' = K$, $J' = I$, $K' = J$ for the first two recursive steps. This gives an algorithm of rank M^3 for the square MM problem, $\langle IJK, IJK, IJK\rangle$. ∎

The proof of Theorem 2.1 is constructive. The coefficients of the resulting bilinear algorithm for $\langle N, N, N\rangle$ can be explicitly expressed through the coefficient of the original one given for $\langle I, J, K\rangle$.

Remark 2.1. More precisely, the tensor of the coefficients of the resulting algorithm is the tensorial power of the tensor of the coefficients of the original algorithm if $I = J = K$. If I, J, K are arbitrary, the former tensor is the tensorial power of the tensor of the algorithm

for $\langle IJK, IJK, IJK\rangle$. The latter tensor is the product of the three tensors of the three dual algorithms that include the original one. We will not use this easily verified fact but we will apply the name *Tensor Product Construction (TPC)* to the recursive process of the proof of Theorem 2.1.

Theorem 2.1 leads to the problem of the design of bilinear algorithms for $\langle I, J, K\rangle$ where $\log M/\log(IJK)$ is as small as possible. Before involving ourselves with that main problem we would like to warn the reader that we do not mean to define the smallest exponent of MM in this way. To be formal, we introduce the following definition which will also be used in the next sections.

Definition 2.3. Let a bilinear arithmetic computational problem be defined by a set of bilinear forms $\mathcal{B}$, or by a trilinear form $T(\underline{X}, \underline{Y}, \underline{Z})$, or by its tensor t. Then $\rho(\mathcal{B}) = \rho(T) = \rho(t)$, the *rank of the problem, of its tensor* t, and *of the trilinear form* $T(\underline{X}, \underline{Y}, \underline{Z})$ is the minimum rank of all bilinear algorithms that solve this problem. For arbitrary natural numbers I, J, K, the rank of $\langle I, J, K\rangle$ is designated by $\rho(\langle I, J, K\rangle)$. (The rank may depend on the choice of the field of constants F so that strictly speaking we have to write ρ_F rather than ρ. Usually we will omit the subscript F assuming that F is fixed; see also Theorem 2.3 below.)

Using the tensor product construction we obtain $\big(\rho(\langle I, J, K\rangle)\big)^h \geq \rho(\langle I^h, J^h, K^h\rangle)$ for all natural h. On the other hand, it is known (see [24, 25]) that

$$\rho(\langle I, J, 1\rangle) = IJ\,,\ \rho(\langle I, J, K\rangle) \geq (I-1)(J+1) + JK \quad \text{if } K > 1\,. \tag{2.10}$$

In particular, $\rho(\langle 2, 2, 2\rangle) \geq 7$ and in fact, $\rho(\langle 2, 2, 2\rangle) = 7$, see [1]. If we choose $I = J = K = 2$ and apply Theorem 2.1 then we only obtain the estimate $\rho(\langle 2^h, 2^h, 2^h\rangle) \leq 7^h$ while it is known that $\rho(\langle 2^h, 2^h, 2^h\rangle) < 7^h$ for all $h \geq 5$; see [9]. Combining the two techniques based on the concept of APA-algorithms (see [5, 6]) and on the 2-Procedure of trilinear aggregating (see [3, 4, 9, 12]) it is easy to prove more general results of this kind; see [17] and compare [13].

Theorem 2.2. *For arbitrary I, J, K, $\rho(\langle I, J, K\rangle)^h > \rho(\langle I^h, J^h, K^h\rangle)$ for all sufficiently large h.*

Notice that Theorem 2.2 does not lead to any improvement of the lower bounds (2.10). The meaning of Theorem 2.2 is that any given exponent of MM associated with constant $c = 1$ can be further reduced. It is not clear if there exists the minimum exponent of MM. ($\beta = 2$ could be a candidate.) However certainly the asymptotic arithmetic complexity of MM can be represented by $\beta^*_{min} = \beta^*_{min}(F)$, the *smallest limiting exponent* of MM which is, of course, unique if the field of constants F is given. Moreover, it is easy to prove a stronger statement on the uniqueness.

Theorem 2.3. *The smallest limiting exponent of MM over F does not depend on the choice of an infinite field of constants F so that for any infinite field F*

$$\beta^*_{\min}(F) = \beta^*_{\min}(Q) = \beta^*_{\min}(C)$$

where Q, C are the fields of rational and complex numbers respectively.

Proof. It is known that any infinite field is isomorphic to an infinite subfield of C (and such a subfield always contains Q). Thus we can assume that all constants from F are complex numbers. Then for arbitrary $\epsilon > 0$ there exist integers $I = I(\epsilon)$, $J = J(\epsilon)$, $K = K(\epsilon)$ such that

$$\log \rho_F(\langle I, J, K\rangle)/\log(IJK) < \beta^*_{\min}(F) + \epsilon . \tag{2.11}$$

As is easy to verify (see [4]), the existence of a bilinear algorithm for $\langle I, J, K\rangle$ of a fixed rank M, in particular of the rank $M = \rho_F(\langle I, J, K\rangle)$ is equivalent to the existence of a solution of a system of algebraic equations with coefficients 0 and 1. It follows that

$$\rho_F(\langle I, J, K\rangle) = \rho_E(\langle I, J, K\rangle) \tag{2.12}$$

where $E = E(Q)$ is an algebraic extension of Q. (2.11) and (2.12) imply that $\beta^*_{\min}(F) + \epsilon$ is an exponent of MM over E so that

$$\beta^*_{\min}(E) \le \beta^*_{\min}(F) + \epsilon . \tag{2.13}$$

Theorem 2.3 follows from (2.13) for $\epsilon \to 0$ if we recall that

$$\beta^*_{\min}(E) = \beta^*_{\min}(Q);$$

see, for instance, [9, Theorem 3.2]. ∎

Throughout the paper our results do not depend on the choice of F unless it is stated otherwise.

3. The Direct Sum of Problems and the Direct Sum Problem. Tensor Product Construction for Direct Sums.

In this section we generalize Theorem 2.1 and apply it to the case where several independent matrix products are to be evaluated. We will define this problem as a particular case of direct sum of r bilinear problems.

Definition 3.1. Given a field F of constants and r sets of bilinear forms $\mathcal{B}^{(1)}, \ldots, \mathcal{B}^{(r)}$ such that

$$\mathcal{B}^{(s)} = \{B^{(s)}_\eta(\underline{X}^{(s)}, \underline{Y}^{(s)})\}, \quad s = 1, \ldots, r, \tag{3.1}$$

$$\underline{X} = (\underline{X}^{(1)}, \ldots, \underline{X}^{(r)}), \quad \underline{Y} = (\underline{Y}^{(1)}, \ldots, \underline{Y}^{(r)}), \tag{3.2}$$

and the entries of the vectors $\underline{X}$, $\underline{Y}$ are indeterminates. (The latter condition implies that the sets $\mathcal{B}^{(1)}, \ldots, \mathcal{B}^{(r)}$ are disjoint, that is, the sets of their input variables are independent each of others.) The problem of simultaneous evaluation of the set $\{\mathcal{B}^{(1)}, \ldots, \mathcal{B}^{(r)}\}$ over F is called the *direct sum of the r bilinear problems* $\mathcal{B}^{(1)}, \ldots, \mathcal{B}^{(r)}$ and is designated by

$$\mathcal{B} = \sum_{s=1}^{r} \oplus \mathcal{B}^{(s)}. \tag{3.3}$$

In particular, if r products $\underline{X}^{(s)}\underline{Y}^{(s)}$ of $I(s) \times J(s)$ by $J(s) \times K(s)$ matrices $X^{(s)}$ and $Y^{(s)}$ respectively are to be evaluated over F for $s = 1, \ldots, r$ and the entries of all matrices $X^{(s)}$, $Y^{(s)}$ are indeterminates then such a *direct sum of r problems of* MM is designated by $\sum_{s=1}^{r} \oplus\langle I(s), J(s), K(s)\rangle$.

The direct sum of r problems, $\mathcal{B}$ (see (3.1)-(3.3)) can be equivalently represented by the set $\{\mathcal{B}^{(1)}, \ldots, \mathcal{B}^{(r)}\}$ of bilinear forms, by the tensor of their coefficients, and by the following trilinear form,

$$T(\underline{X}, \underline{Y}, \underline{Z}) = \sum_{s} T^{(s)}(\underline{X}^{(s)}, \underline{Y}^{(s)}, \underline{Z}^{(s)}), \tag{3.4}$$

$$T^{(s)}(\underline{X}^{(s)}, \underline{Y}^{(s)}, \underline{Z}^{(s)}) = \sum_{\eta} B_{\eta}^{(s)}(\underline{X}^{(s)}, \underline{Y}^{(s)}) z_{\eta}^{(s)}. \tag{3.5}$$

Here

$$\underline{Z}^{(s)} = (z_{\eta}^{(s)}), \quad \underline{Z} = (\underline{Z}^{(1)}, \ldots, \underline{Z}^{(s)}) \tag{3.6}$$

are vectors of indeterminates, $z_{\eta}^{(s)}$ and $T^{(s)}(\underline{X}^{(s)}, \underline{Y}^{(s)}, \underline{Z}^{(s)})$ are trilinear problems that define the bilinear problems $\mathcal{B}^{(s)}$; see (3.1)-(3.3).

In the case of $\sum_{s=1}^{r} \oplus\langle I(s), J(s), K(s)\rangle$,

$$T(\underline{X}, \underline{Y}, \underline{Z}) = \sum_{s=1}^{r} \mathrm{Tr}(X^{(s)} Y^{(s)} Z^{(s)}) \tag{3.7}$$

where $Z^{(s)}$ is the $K(s) \times I(s)$ matrix whose entries are indeterminates, $s = 1, \ldots, r$.

As is obvious, the solution of an arbitrary direct sum of r bilinear problems can be obtained if each of the r problems is solved independently of other $r - 1$ ones. Such a solution is represented by the following r decompositions,

$$T^{(s)}(\underline{X}^{(s)}, \underline{Y}^{(s)}, \underline{Z}^{(s)}) = \sum_{q=1}^{M(s)} L_{qs}(\underline{X}^{(s)}) L'_{qs}(\underline{Y}^{(s)}) L''_{qs}(\underline{Z}^{(s)}) \quad \text{for } s = 1, \ldots, r. \tag{3.8}$$

An algorithm defined by (3.8) is called a *direct sum algorithm* and has rank $M = \sum_{s=1}^{r} M(s)$. However we might hope to take advantage by solving the r problems simultaneously. Such a solution is defined by the more general decomposition, (2.6) and consequently gives (bilinear) algorithm of a more general class.

In the case of direct sums of several bilinear problems, the $L_q(\underline{X})$, $L'_q(\underline{Y})$, $L''_q(\underline{Z})$ in (2.6) can be defined by the following identities (rather than by (2.8), (2.9)).

$$\forall q: \quad L_q(\underline{X}) = \sum_{\mu,s} f_{q\mu s} x_{\mu}^{(s)}, \quad L'_q(\underline{Y}) = \sum_{\nu,s} f'_{q\nu s} y_{\nu}^{(s)}, \quad L''_q(\underline{Z}) = \sum_{\eta,s} f''_{q\eta s} z_{\eta}^{(s)}, \tag{3.9}$$

$$f_{q\mu s}, f'_{q\nu s}, f''_{q\eta s} \in F \quad \text{for all } q, \mu, \nu, \eta, s. \tag{3.10}$$

(On the other hand, (3.9), (3.10) can be represented as a particular case of (2.8), (2.9).)

Again in the case of MM, μ, ν, η are defined by the pairs (i, j), (j, k), and (k, i) respectively. Notice that the μ, ν, η (and in the case of MM also the i, j, k) range in the domains that depend on s.

Now the problem arises if there exist algorithms (2.6), (3.9), (3.10) that are indeed faster than the best direct sum algorithms (3.8)? In particular, does there exist r disjoint bilinear problems $\mathcal{B}^{(1)},\ldots,\mathcal{B}^{(r)}$ such that

$$\rho\Big(\sum_{s=1}^{r}\oplus\mathcal{B}^{(s)}\Big) < \sum_{s=1}^{r}\rho(\mathcal{B}^{(s)})\,? \tag{3.11}$$

The latter problem is called the *Direct Sum Problem (DSP)*. The *Direct Sum Conjecture (DSC)* is that (3.11) never holds. We are interested in the DSP and DSC for the class of MM algorithms.

Let us assume for a while that the DSC for MM is true. Then Theorem 2.1 can be generalized in the following straightforward manner.

Proposition 3.1. Given a bilinear algorithm (2.6), (3.9), (3.10) of rank M for the direct sum of r disjoint problems of MM, $\langle I(s),J(s),K(s)\rangle$, $s=1,\ldots,r$ where M, r, $I(s)$, $J(s)$, $K(s)$ for $s=1,\ldots,r$ are arbitrary, $M>r$. Let $\tau=\tau^*$ be the real solution to the following equation,

$$\sum_{s=1}^{r}\big(I(s)J(s)K(s)\big)^{\tau} = M\,. \tag{3.12}$$

Then the DSC implies that $\beta^*=3\tau^*$ is an exponent of MM.

Definition 3.2. The equation (3.12) is called the *equation associated with a bilinear algorithm of rank M for $\sum_{s=1}^{r}\oplus\langle I(s),J(s),K(s)\rangle$*.

Proof. Let real $\tau(s)$ satisfy the following equations

$$\rho(\langle Is),J(s),K(s)\rangle) = \big(I(s)J(s)K(s)\big)^{\tau(s)} \tag{3.13}$$

where $s=1,\ldots,r$. Using the DSC we obtain

$$\sum_{s=1}^{r}\rho\langle I(s),J(s),K(s)\rangle = \rho\Big(\sum_{s=1}^{r}\oplus\langle I(s),J(s),K(s)\rangle\Big) \le M\,. \tag{3.14}$$

Combining (3.13) and (3.14) gives

$$M \ge \sum_{s=1}^{r}\big(I(s)J(s)K(s)\big)^{\tau(s)} \ge \sum_{s=1}^{r}\big(I(s)J(s)K(s)\big)^{\tau_{\min}} \tag{3.15}$$

where $\tau_{\min}=\min_s\tau(s)$. By virtue of Theorem 2.1, $3\tau(s)$ for all s and hence $3\tau_{\min}$ are exponents of MM. Comparing (3.12) and (3.15) gives $\tau_{\min}\le\tau^*$. ∎

Proposition 3.1 motivates Definition 3.2, but we could apply that Proposition only if the DSC is proven to be true for MM. This is still an open problem (see the Introduction).

Fortunately a generalization of the Tensor Product Construction (TPC) enables us to save the most essential part of the result of Proposition 3.1.

Theorem 3.1 (Exponential Direct Sum Theorem, EDST). *Under the conditions of Proposition 3.1, the $\beta^*=3\tau^*$ is a limiting exponent of MM (even if the DSC is false).*

To prove Theorem 3.1 we first generalize the TPC.

Hereafter we designate

$$r \odot \langle I, J, K \rangle = \sum_{s=1}^{r} \oplus \langle I, J, K \rangle, \quad r\, r' \odot \langle I, J, K \rangle = r \odot r' \odot \langle I, J, K \rangle \tag{3.16}$$

for arbitrary natural r, r', I, J, K.

Using this notation we represent a bilinear algorithm (2.6), (3.4) (3.7) as the following mapping,

$$\sum_{s=1}^{r} \oplus \langle I(s), J(s), K(s) \rangle \leftarrow M \odot \langle 1, 1, 1 \rangle . \tag{3.17}$$

The right side of (3.17) represents the right side of (2.6) where each product $L_q(\underline{X}) L'_q(\underline{Y}) L''_q(\underline{Z})$ is represented as $\langle 1, 1, 1 \rangle$.

We recall the basic observation of the proof of Theorem 2.1 (which has led us to the TPC) that the substitution of $I \times J$, $J \times K$ and $K \times I$ matrices for the entries of $\underline{X}$, $\underline{Y}$, $\underline{Z}$ respectively preserves (2.6). Such a substitution turns the direct sum of the left side of (3.17) into the direct sum $\sum_{s=1}^{r} \oplus \langle I(s)I, J(s)J, K(s)K \rangle$. Also it turns each of the products $L_q(\underline{X}) L'_q(\underline{Y}) L''_q(\underline{Z})$ into the product of $I \times J$ by $J \times K$ by $K \times I$ matrices. Hence the substitution turns (3.17) into an algorithm that can be represented by the following mapping,

$$\sum_{s=1}^{r} \oplus \langle I(s)I, J(s)J, K(s)K \rangle \leftarrow M \odot \langle I, J, K \rangle . \tag{3.18}$$

We will state the latter result formally as Lemma 3.2 using the following definition.

Definition 3.3. A *mapping* $\mathcal{B} \leftarrow \mathcal{B}'$ *is valid* if there exists a bilinear algorithm that is represented by such a mapping. Then we write $\mathcal{B} \leftarrow\!: \mathcal{B}'$. (In this paper we use the notation $\mathcal{B} \leftarrow\!: \mathcal{B}'$ mostly in the cases where $\mathcal{B}' = M \odot \langle 1, 1, 1 \rangle$.)

Lemma 3.2. *If (3.17) is valid then (3.18) is valid.*

Equation (3.18) can be interpreted as the product of (3.17) by the trivial mapping

$$\langle I, J, K \rangle \leftarrow \langle I, J, K \rangle \tag{3.19}$$

for aribtrary natural I, J, K.

Similarly we can define the valid trivial mapping

$$\sum_{\ell=1}^{r'} \oplus \langle I'_\ell, J'_\ell, K'_\ell \rangle \leftarrow \sum_{\ell=1}^{r'} \oplus \langle I'_\ell, J'_\ell, K'_\ell \rangle \tag{3.20}$$

for arbitrary natural r', I'_ℓ, J'_ℓ, K'_ℓ, $\ell = 1, \ldots, r'$.

Multiplying (3.17) and (3.20) we obtain the following mapping,

$$\sum_{s=1}^{r} \sum_{\ell=1}^{r'} \oplus \langle I_s I'_\ell, J_s J'_\ell, K_s K'_\ell \rangle \leftarrow M \odot \sum_{\ell=1}^{r'} \oplus \langle I'_\ell, J'_\ell, K'_\ell \rangle . \tag{3.21}$$

The meaning of the direct sum in the left side is obvious. The M terms of the direct sum in the right side of (3.21) represent the M sets each consisting of r' products $l_{q\ell}(\underline{X}^{(\ell)})l'_{q\ell}(\underline{Y}^{(\ell)})l''_{q\ell}(\underline{Z}^{(\ell)})$, $\ell = 1, \ldots, r'$, $q = 1, \ldots, M$, where $\underline{X}^{(\ell)}$, $\underline{Y}^{(\ell)}$, $\underline{Z}^{(\ell)}$ are $I'_\ell \times J'_\ell$, $J'_\ell \times K'_\ell$, and $K'_\ell \times I'_\ell$ matrices respectively.

To justify the validity of (3.21) (assuming the validity of (3.17)), apply Lemma 3.2 for $I = I'_\ell$, $J = J'_\ell$, $K = K'_\ell$, $\ell = 1, \ldots, r'$. Then apply the following simple lemma.

Lemma 3.3. *$B \leftarrow: B'$ and $\hat{B} \leftarrow: \hat{B}'$ imply*

$$\hat{B} \oplus B \leftarrow: B' \oplus \hat{B}'.$$

We have proven the following generalization of Lemma 3.2 and of the basic observation for the tensor product construction.

Lemma 3.4. *If mapping (3.17) is valid then mapping (3.21) is valid.*

We also need the two following simple lemmas.

Lemma 3.5. *$B \leftarrow: B'$ and $B' \leftarrow: B''$ imply $B \leftarrow: B''$.*

Lemma 3.6. *The mapping*

$$\ell'(q) \odot (I(q), J(q), K(q)) \leftarrow \sum_{s=1}^{r} \ell(s) \odot (I(s), J(s), K(s))$$

is valid for arbitrary natural q, $\ell'(q)$, r, $\ell(s)$, $I(s)$, $J(s)$, $K(s)$, $s = 1, \ldots, r$ if $1 \le q \le r$, $\ell'(q) \le \ell(q)$.

Now we are ready to prove the main lemma of this section.

Lemma 3.7. *Given arbitrary natural numbers ℓ, I, J, K, r. Let*

$$r \odot (I, J, K) \leftarrow: \ell r \odot (1, 1, 1). \tag{3.22}$$

Then the mappings

$$r \odot (I^h, J^h, K^h) \leftarrow \ell^h r \odot (1, 1, 1) \tag{3.23}$$

are valid for $h = 1, 2, 3, \ldots$.

Proof (by induction in h). Let (3.23) be valid for $h = h^*$. Then by virtue of Lemma 3.2,

$$r \odot (I^{h^*+1}, J^{h^*+1}, K^{h^*+1}) \leftarrow: \ell^{h^*} \odot (r \odot (I, J, K)). \tag{3.24}$$

(See the notation of (3.16).) Applying Lemma 3.5 to (3.22) and (3.24) we obtain that (3.23) is valid for $h = h^* + 1$. Observe that (3.23) for $h = 1$ is the given valid mapping (3.22). ∎

Next we restate Theorem 3.1 in the following obviously equivalent form and then prove it.

Theorem 3.1. *Let for some natural numbers M, r, $I(s)$, $J(s)$, $K(s)$, $s = 1,\dots,r$, $r < M$, the mapping (3.17) be valid and $\tau = \tau^*$ be the real solution of the associated equation (3.12). Then $\beta^* = 3\tau^*$ is a limiting exponent of MM.*

Proof. Observe that Theorem 2.1 and Lemmas 3.5–3.7 imply Theorem 3.1 in the case where the valid basic mapping (3.17) takes the form (3.22). (Indeed, consider valid mapping (3.23) where h is sufficiently large, apply Lemmas 3.5, 3.6 in order to delete r in the left side, and then apply Theorem 2.1.)

Finally consider the general case of arbitrary valid basic mapping (3.17). Recursively applying Lemmas 3.4, 3.5 to (3.17) we obtain the following sequence of valid mappings for $h = 1, 2, 3, \dots$,

$$\sum_{\underline{\alpha}\in Q(h,r)} \oplus c(\underline{\alpha}) \odot \langle I(\underline{\alpha}), J(\underline{\alpha}), K(\underline{\alpha})\rangle \leftarrow M^h \odot \langle 1,1,1\rangle . \tag{3.25}$$

Here $Q(h,r)$ is the set of r-dimensional vectors $\underline{\alpha} = (\alpha_1,\dots,\alpha_r)$ with nonnegative integer entries $\alpha_1,\dots,\alpha_r$ such that

$$\alpha_1 + \dots + \alpha_r = h, \tag{3.26}$$

$$c(\underline{\alpha}) = \frac{h!}{\alpha_1!\,\alpha_2!\,\dots\,\alpha_r!} \le r^h, \tag{3.27}$$

$$I(\underline{\alpha}) = \prod_{s=1}^{r} (I(s))^{\alpha_s},\ J(\underline{\alpha}) = \prod_{s=1}^{r} (J(s))^{\alpha_s},\ K(\underline{\alpha}) = \prod_{s=1}^{r} (K(s))^{\alpha_s}. \tag{3.28}$$

Mapping (3.25) (3.28) can be considered the h-th power of (3.17). We used the well known formula of multinomial expansion to represent the terms in the left side of (3.25). The mapping (3.17) coincides with the mapping (3.25) (3.28) for $h = 1$.

Simultaneously with the sequence of mappings (3.25) (3.28) we define the following sequence of the associated equations in τ.

$$\sum_{\underline{\alpha}\in Q(h,r)} c(\underline{\alpha})\big(I(\underline{\alpha})J(\underline{\alpha})K(\underline{\alpha})\big)^{\tau} = M^h, \quad h = 1,2,3,\dots . \tag{3.29}$$

We observe that for all τ and for all h

$$\sum_{\underline{\alpha}\in Q(h,r)} c(\underline{\alpha})\big(I(\underline{\alpha})\,J(\underline{\alpha})\,K(\underline{\alpha})\big)^{\tau} = \Big(\sum_{s=1}^{r}\big(I(s)J(s)K(s)\big)^{\tau}\Big)^{h}.$$

It follows that the equations (3.29) have the same (real) solution for all h which coincide with the solution $\tau = \tau^*$ of the equation (3.12).

Let $\underline{\alpha}^*(h)$ be a vector from $Q(h,r)$ such that

$$c(\underline{\alpha}^*)\big(I(\underline{\alpha}^*)J(\underline{\alpha}^*)K(\underline{\alpha}^*)\big)^{\tau^*} = \max_{\underline{\alpha}\in Q(h,r)} c(\underline{\alpha})\big(I(\underline{\alpha})\,J(\underline{\alpha})\,K(\underline{\alpha})\big)^{\tau^*} \ge M^h/|Q(h,r)| \tag{3.30}$$

where $|Q(h,r)| = (r+h)!/h!$ is the cardinality of the set $Q(h,r)$.

As follows from the validity of mapping (3.25)–(3.28) and from Lemmas 3.5 and 3.6,

$$c \odot \langle I(\underline{\alpha}^*), J(\underline{\alpha}^*), K(\underline{\alpha}^*)\rangle \leftarrow: M^h \odot \langle 1,1,1\rangle$$

for all $c \le c(\underline{\alpha}^*)$. We choose $c = M^g$ where g is the natural number such that $M^g \le c(\underline{\alpha}^*) < M^{g+1}$. Then we come to a valid mapping which can be represented in the form (3.22). Hence the real solution $\tau = \tau(h)$ to the associated equation

$$M^g\big(I(\underline{\alpha}^*)\,J(\underline{\alpha}^*)\,K(\underline{\alpha}^*)\big)^{\tau} = M^h \tag{3.31}$$

is a limiting exponent of MM.

On the other hand, since the cardinality of $Q(h,r)$ is equal to $(r+h)!/h!$, (3.29), (3.30) imply the next relations,

$$c(\underline{\alpha}^*)\big(I(\underline{\alpha}^*)\,J(\underline{\alpha}^*)\,K(\underline{\alpha}^*)\big)^{\tau^*} \ge \frac{h!}{(r+h)!}\sum_{\underline{\alpha}\in Q(h,r)} c(\underline{\alpha})\big(I(\underline{\alpha})\,J(\underline{\alpha})\,K(\underline{\alpha})\big)^{\tau^*} = M^h\,\frac{h!}{(r+h)!}\,.$$

Since $M^g > c(\underline{\alpha}^*)/M$ and since $\big(I(\underline{\alpha}^*)J(\underline{\alpha}^*)K(\underline{\alpha}^*)\big)^{\epsilon} > \frac{M(r+h)!}{h!}$ for all $\epsilon > 0$ and for all sufficiently large h (see (3.27), (3.30) and recall that $M > r$), it follows that for arbitrary $\epsilon > 0$

$$M^g\big(I(\underline{\alpha}^*)\,J(\underline{\alpha}^*)\,K(\underline{\alpha}^*)\big)^{\tau^*+\epsilon} > M^h \tag{3.32}$$

if $h = h(\epsilon)$ is chosen sufficiently large.

Comparing (3.31) for $\tau = \tau(h)$ and (3.32) we obtain for arbitrary $\epsilon > 0$

$$\tau^* + \epsilon > \tau\big(h(\epsilon)\big)\,.$$

Hence $\tau^* + \epsilon$ is an exponent of MM for any $\epsilon > 0$. ∎

4. Trilinear Aggregating Generated by Tables.

In this section we introduce the techniques of trilinear aggregating, TA, in new modified versions and describe the method in a more formal and more general way than we did earlier. We start with an illustrative example of TA.

Example 4.1. (2-Procedure.)

$$\mathrm{Tr}(XYZ) + \mathrm{Tr}(UVW) = \sum_{i,j,k}(x_{ij}+u_{jk})(y_{jk}+v_{ki})(z_{ki}+w_{ij}) - \sum_{i,j}x_{ij}\sum_{k}(y_{jk}+v_{ki})w_{ij}$$
$$- \sum_{j,k}u_{jk}y_{jk}\sum_{i}(z_{ki}+w_{ij}) - \sum_{k,i}\Big(\sum_{j}(x_{ij}+u_{jk})\Big)v_{ki}z_{ki}\,.$$

To simplify the formula we have slightly deviated from our previous notation writing X, Y, Z, U, V, W rather than $X^{(1)}$, $Y^{(1)}$, $Z^{(1)}$, $X^{(2)}$, $Y^{(2)}$, $Z^{(2)}$ respectively. Let X, Y, Z, U, V, W be $I\times J$, $J\times K$, $K\times I$, $J\times K$, $K\times I$, $I\times J$ matrices respectively and let i, j, k in the above identities range from 0 to $I-1$, $J-1$ and $K-1$ respectively. Then the 2-Procedure implies that for arbitrary natural I, J, K:

$$\langle I,J,K\rangle \oplus \langle J,K,I\rangle \leftarrow: (IJK + IJ + JK + KI) \odot \langle 1,1,1\rangle\,.$$

The 2-Procedure of TA can be deduced from the following table.

Table 4.1.

x_{ij}	y_{jk}	z_{ki}
u_{jk}	v_{ki}	w_{ij}

We will explain how to define TA by the following more general tables.

Table 4.2.

$x^{(1)}_{i(1)j(1)}$	$y^{(1)}_{j(1)k(1)}$	$z^{(1)}_{k(1)i(1)}$
$x^{(2)}_{i(2)j(2)}$	$y^{(2)}_{j(2)k(2)}$	$z^{(2)}_{k(2)i(2)}$
...	...	...
$x^{(r)}_{i(r)j(r)}$	$y^{(r)}_{j(r)k(r)}$	$z^{(r)}_{k(r)i(r)}$

Definition 4.1. Given an $r \times 3$ table (Table 4.2) whose entries $(s,1)$, $(s,2)$, $(s,3)$ are filled with the indeterminates $x^{(s)}_{i(s)j(s)}$, $y^{(s)}_{j(s)k(s)}$, $z^{(s)}_{k(s)i(s)}$, respectively. Then the table is called *Generating Table, GT*. The product

$$\pi(q,s,t) = x^{(q)}_{i(q)j(q)}\, y^{(s)}_{j(s)k(s)}\, z^{(t)}_{k(t)i(t)} \tag{4.1}$$

is called either the *s-th principal term of the GT* if $q = s = t$ or the *correction term* (q,s,t) *of the* GT otherwise. The product $\sum_{q=1}^{r} x^{(q)}_{i(q)j(q)} \sum_{s=1}^{r} y^{(s)}_{j(s)k(s)} \sum_{t=1}^{r} z^{(t)}_{k(t)i(t)}$ is called *the aggregate* of the table.

Table 4.1 is an example of GT where $r = 2$, $i(1) = i$, $j(1) = j$, $k(1) = k$, $i(2) = j$, $j(2) = k$, $k(2) = i$.

The next result is easy to verify.

Lemma 4.1. *Given Generating Table 4.2 then its aggregate is identically the sum of all its principal and correction terms.*

Hereafter we assume that the $3r$ subscripts $i(s)$, $j(s)$, $k(s)$, $s = 1, \ldots, r$ in the GT are integer variables that independently of each other range from 0 to some fixed bounds $I-1$, $J-1$, $K-1$. We designate that

$$H = IJK\,. \tag{4.2}$$

Remark 4.1. We will not use the obvious possibility to generalize our construction to the case where $I = I(s)$, $J = J(s)$, $K = K(s)$ depend on s but $H = I(s)J(s)K(s)$ does not depend on s.

Then there exist H instances of such a GT and therefore H instances of each principal term, of each correction term, and of the aggregate of that GT. The next simple fact is important for us.

Lemma 4.2. *The sum of the H instances of the s-th principal terms of Generating Table 4.2 is identically the $Tr(X^{(s)}Y^{(s)}Z^{(s)})$ where $X^{(s)} = (x^{(s)}_{i(s)j(s)})$, $Y^{(s)} = (y^{(s)}_{j(s)k(s)})$, $Z^{(s)} = (z^{(s)}_{k(s)i(s)})$, are $I \times J$, $J \times K$, $K \times I$ matrices respectively.*

Corollary 4.1. *Given H instances of Generating Table 4.2 (where (4.2) holds). Then*

$$r \odot (I, J, K) \leftarrow: (H + \rho c) \odot (1, 1, 1) \tag{4.3}$$

where ρc is the rank of the sum of the H instances of all correction terms of the GT. (See Definition 2.3 about the ranks of trilinear forms.)

Indeed, the sum of the H instances of the aggregates gives $H \odot (1, 1, 1)$. Subtracting the sum of all instances of all correction terms gives(4.3) by virtue of Lemmas 4.1 and 4.2. ∎

In the sequel we combine Corollary 4.1 with the techniques of Implicit Canceling of correction terms of Table 4.2, see Section 7.

5. Generating Tables with Vectors as Subscripts.

In this section we combine the TPC and TA. Let m, n be natural numbers. Consider the m-dimensional vector $\underline{h} = (h(1), \ldots, h(m))$ where $h(g)$ are independent integer parameters that range from 0 to $n-1$. Consider also r different partitions of the vector $\underline{h}$ into $\underline{i}(s)$, $\underline{j}(s)$, $\underline{k}(s)$, $s = 1, \ldots, r$, its three disjoint subvectors of dimensions ℓ, ℓ', ℓ'' respectively where ℓ, ℓ', ℓ'', r are fixed natural numbers such that

$$\ell + \ell' + \ell'' = m, \quad r \le m!/(\ell!\,\ell'!\,\ell''!). \tag{5.1}$$

Remark 5.1. Here and hereafter we assume that the order of the entries of a vector is preserved for its subvectors.

We will use the following notation to represent the s-th partition of the vector $\underline{h}$,

$$\underline{h} = \underline{i}(s)\underline{j}(s)\underline{k}(s) \qquad \text{for } s = 1, \ldots, r. \tag{5.2}$$

$$\underline{i}(s) = \big(i(1,s), \ldots, i(\ell,s)\big), \quad i(t,s) = h\big(q(t,s)\big), \quad t = 1, \ldots, \ell, \tag{5.3}$$

$$\underline{j}(s) = \big(j(1,s), \ldots, j(\ell',s)\big), \quad j(t',s) = h\big(q'(t',s)\big), \quad t' = 1, \ldots, \ell', \tag{5.4}$$

$$\underline{k}(s) = \big(k(1,s), \ldots, k(\ell'',s)\big), \quad k(t'',s) = h\big(q''(t'',s)\big), \quad t'' = 1, \ldots, \ell'', \tag{5.5}$$

Since for all s the entries of $\underline{i}(s)$, $\underline{j}(s)$, $\underline{k}(s)$ coincide with some entries of $\underline{h}$, they are also integer parameters that range from 0 to $n-1$.

Now we establish the following obvious one-to-one correspondence between the triplets of vectors $\big(\underline{i}(s), \underline{j}(s), \underline{k}(s)\big)$ and integers $\big(i(s), j(s), k(s)\big)$,

$$i(s) = \sum_{t=1}^{\ell} i(t,s)n^{t-1}, \; j(s) = \sum_{t'=1}^{\ell'} j(t',s)n^{t'-1}, \; k(s) = \sum_{t''=1}^{\ell''} k(t'',s)n^{t''-1}. \tag{5.6}$$

This implies that $i(s)$, $j(s)$, $k(s)$ range from 0 to $I-1$, $J-1$, $K-1$ respectively where

$$I = n^{\ell}, \quad J = n^{\ell'}, \quad K = n^{\ell''}, \quad IJK = n^m = H. \tag{5.7}$$

(Compare (4.2).)

Now we can rewrite Generating Table 4.2 in the following equivalent form.

Table 5.1.

$$\begin{array}{lll} x^{(1)}_{\underline{i}(1)\underline{j}(1)} & y^{(1)}_{\underline{j}(1)\underline{k}(1)} & z^{(1)}_{\underline{k}(1)\underline{i}(1)} \\ x^{(2)}_{\underline{i}(2)\underline{j}(2)} & y^{(2)}_{\underline{j}(2)\underline{k}(2)} & z^{(2)}_{\underline{k}(2)\underline{i}(2)} \\ \cdots & \cdots & \cdots \\ x^{(r)}_{\underline{i}(r)\underline{j}(r)} & y^{(r)}_{\underline{j}(r)\underline{k}(r)} & z^{(r)}_{\underline{k}(r)\underline{i}(r)} \end{array}$$

We will consider Tables 4.2 and 5.1 identical assuming that

$$x^{(s)}_{\underline{i}(s)\underline{j}(s)} = x^{(s)}_{i(s)j(s)}\,,\ y^{(s)}_{\underline{j}(s)\underline{k}(s)} = y^{(s)}_{j(s)k(s)}\,,\ z^{(s)}_{\underline{k}(s)\underline{i}(s)} = z^{(s)}_{k(s)i(s)}\,. \tag{5.8}$$

(See (5.2)–(5.6).) Consequently we will designate (compare (4.1))

$$\pi(q,s,t) = x^{(q)}_{\underline{i}(q)\underline{j}(q)}\, y^{(s)}_{\underline{j}(s)\underline{k}(s)}\, z^{(t)}_{\underline{k}(t)\underline{i}(t)} \tag{5.9}$$

and also extend the definition of the principal and correction terms and of the aggregate of Table 4.1 as well as Corollary 4.1 to the case of Table 5.1. On the other hand, we will exploit the vector structure of the subscripts of the indeterminates of Table 5.1 in our next definition.

Remark 5.2. Because of the identities (5.8) we will not distinguish between the two bilinear problems associated with Tables 4.2 and 5.1. In particular, we substitute (5.7) in (4.3) and obtain

$$r \odot (n^{\ell}, n^{\ell'}, n^{\ell''}) \leftarrow: (n^m + \rho c) \odot \langle 1,1,1 \rangle\,. \tag{5.10}$$

Definition 5.1. $\deg_g x^{(q)}_{\underline{i}(q)\underline{j}(q)}$ (respectively $\deg_g y^{(s)}_{\underline{j}(s)\underline{k}(s)}$, $\deg_g z^{(t)}_{\underline{k}(t)\underline{i}(t)}$), the *degree* of $x^{(q)}_{\underline{i}(q)\underline{j}(q)}$ (respectively of $y^{(s)}_{\underline{j}(s)\underline{k}(s)}$, of $z^{(t)}_{\underline{k}(t)\underline{i}(t)}$) in $h(g)$ is the number of occurrences of the $h(g)$ among the entries of vectors $\underline{i}(q)$, $\underline{j}(q)$ (respectively $\underline{j}(s)$, $\underline{k}(s)$ or $\underline{k}(t)$, $\underline{i}(t)$) where $1 \le g \le m$, $1 \le q,s,t \le r$. If $\pi(q,s,t)$ (see (5.9)), is *a principal or correction term of Table 5.1 then*

$$\deg_g \pi(q,s,t) = \deg_g x^{(q)}_{\underline{i}(q)\underline{j}(q)} + \deg_g y^{(s)}_{\underline{j}(s)\underline{k}(s)} + \deg_g z^{(t)}_{\underline{k}(t)\underline{i}(t)}\,. \tag{5.11}$$

$\pi(q,s,t)$ is a *product of degree* 1 if it has degree 1 in $h(g)$ for at least one value g, $1 \le g \le m$. The *dimension of* $\pi(q,s,t)$ is the number of different g such that the degree of the $\pi(q,s,t)$ in the $h(g)$ is positive.

The next simple estimates follow from the fact that all of the entries of the three vectors $\underline{i}(s)$, $\underline{j}(s)$, $\underline{k}(s)$ are different parameters.

Lemma 5.1. *Each principal term of Table 5.1 has degree 2 in all $h(g)$, $g = 1,\ldots,m$. The degree of each correction term of Table 5.1 in any $h(g)$ is at most 3.*

The next result follows from Definition 5.1 (see in particular (5.11)). It is important for our designs of fast MM algorithms in the next sections.

Lemma 5.2. *Let $\pi(q,s,t)$ (see (5.9)), a correction term of Table 5.1 have degree 1 in $h(g)$ for some g, q, s, t, $1 \le g \le m$, $1 \le q,s,t \le r$. Then the sum*

$$\beta_g(q,s,t) = \sum_{h(g)=0}^{n-1} \pi(q,s,t) \tag{5.12}$$

has rank 1 and, more specifically,

$$\beta_g(q,s,t) = \Big(\sum_{h(g)=0}^{n-1} x^{(q)}_{\underline{i}(q)\underline{j}(q)}\Big) y^{(s)}_{\underline{j}(s)\underline{k}(s)}\, z^{(t)}_{\underline{k}(t)\underline{i}(t)} \quad \text{if } deg_g\, x^{(q)}_{\underline{i}(q)\underline{j}(q)} = 1\,, \tag{5.13}$$

$$\beta_g(q,s,t) = x^{(q)}_{\underline{i}(q)\underline{j}(q)}\Big(\sum_{h(g)=0}^{n-1} y^{(s)}_{\underline{j}(s)\underline{k}(s)}\Big) z^{(t)}_{\underline{k}(t)\underline{i}(t)} \quad \text{if } deg_g\, y^{(s)}_{\underline{j}(s)\underline{k}(s)} = 1\,, \tag{5.14}$$

$$\beta_g(q,s,t) = x^{(q)}_{\underline{i}(q)\underline{j}(q)}\, y^{(s)}_{\underline{j}(s)\underline{k}(s)}\Big(\sum_{h(g)=0}^{n-1} z^{(t)}_{\underline{k}(t)\underline{i}(t)}\Big) \quad \text{if } deg_g\, z^{(t)}_{\underline{k}(t)\underline{i}(t)} = 1\,. \tag{5.15}$$

In fact, in Example 4.1 we have already exploited the advantages given by Lemma 5.2 by uniting the correction terms of Table 4.1 into the sum of only $IJ + JK + KL$ products. In Section 7 we will see some additional reasons to seek for Tables 5.1 whose correction terms have degree 1.

6. How to Design Generating Tables with Correction Terms of Degree 1?

In this section we define two constructions of large Generating Tables 5.1 with correction terms of degree 1. In Section 7 we will exploit the latter property. We hope that our constructions will be eventually generalized and improved. We will use the following notation and definition.

Notation 6.1. Λ is the empty (0-dimensional) vector. Let $\underline{\xi}$, $\underline{\Theta}$ be subvectors of a given vector $\underline{h}$. Then $\underline{\xi} \cup \underline{\Theta}$ and $\underline{\xi} \cap \underline{\Theta}$, the two subvectors of $\underline{h}$ are the union and the intersection of $\underline{\xi}$ and $\underline{\Theta}$ respectively. (Then $\underline{\xi} \cup \underline{\Theta} = \underline{\xi\Theta}$ if $\underline{\xi} \cap \underline{\Theta} = \Lambda$; see Remark 5.1 and Equation (5.2).) $h(g)$ is the g-th entry of $\underline{h}$, $h(g) \in \underline{h}$.

Definition 6.1. The *partitions* of two D-dimensional vectors $\underline{\xi}$ and $\underline{\Theta}$ into x disjoint subvectors $\underline{\xi}_1, \ldots, \underline{\xi}_x$ and $\underline{\Theta}_1, \ldots, \underline{\Theta}_x$ are isomorphic if $\xi(g) \in \underline{\xi}_\nu$ implies $\Theta(g) \in \underline{\Theta}_\nu$ for $g = 1, \ldots, D$, $\nu = 1, \ldots, x$.

Now we are ready to describe our First Construction. Let a natural m be a multiple of 4,

$$m = 4u\,, \tag{6.1}$$

and let $\underline{h}_1$, $\underline{h}_2$ be the two $(2u)$-dimensional subvectors of the vector $\underline{h}$ that consist of the first $2u$ and the last $2u$ entries of $\underline{h}$ respectively. Then write

$$r = (2u)!/(u!)^2 . \tag{6.2}$$

Let $\underline{\varphi}(s)$, $\underline{\psi}(s)$ for $s = 1, \ldots, r$ partition $\underline{h}_1$ into pairs of disjoint u-dimensional subvectors. Let $\underline{\varphi}'(s)$, $\underline{\psi}'(s)$ for $s = 1, \ldots, r$ be the isomorphic partitions of $\underline{h}_2$. Then we define $\underline{i}(s)$, $\underline{j}(s)$, $\underline{k}(s)$, the vectors-subscripts of Table 5.1 as follows.

$$\underline{i}(s) = \underline{\varphi}(s), \ \underline{j}(s) = \underline{\psi}(s)\underline{\varphi}'(s), \ \underline{k}(s) = \underline{\psi}'(s), \quad s = 1, \ldots, r. \tag{6.3}$$

Now Table 5.1 is defined by the vector $\underline{h}$ and by its r partitions into the triplets of disjoint subvectors $(\underline{i}(s), \underline{j}(s), \underline{k}(s))$ such that (6.1) (6.3) hold. This is our *First Construction*. We call it also the *r-Procedure of TA for $r = (2u)!/(u!)^2$*.

We will use the following result.

Lemma 6.1. *Let Table 5.1 be defined by the r-Procedure of TA for $r = (2u)!/(u!)^2$ where (6.1) (6.3) hold. Then each correction term $\pi(q, s, t)$ of Table 5.1 has degree 1,*

$$\forall q \forall s \forall t \exists g: \ deg_g \pi(q, s, t) = 1 \quad unless \ q = s = t. \tag{6.4}$$

Furthermore for each correction term $\pi(q, s, t)$ of Table 5.1 (see (4.1), (5.9)), and for each g, $1 \leq g \leq m$ either

$$h(g) \in \underline{h}_1, \ deg_g x^{(q)}_{\underline{i}(q)\underline{j}(q)} = 1 \tag{6.5}$$

or

$$h(g) \in \underline{h}_2, \ deg_g y^{(s)}_{\underline{j}(s)\underline{k}(s)} = 1. \tag{6.6}$$

Proof. Equations (6.5), (6.6) immediately follow if one examines the next combination of (5.9) and (6.3),

$$\pi(q, s, t) = x^{(q)}_{\underline{\varphi}(q), \underline{\psi}(q)\underline{\varphi}'(q)} \, y^{(s)}_{\underline{\psi}(s)\underline{\varphi}'(s), \underline{\psi}'(s)} \, z^{(t)}_{\underline{\psi}'(t), \underline{\varphi}(t)}, \ q, s, t = 1, \ldots, r. \tag{6.7}$$

We recall (see Notation 6.1) that

$$\forall s: \ \underline{\varphi}(s)\underline{\psi}(s) = \underline{h}_1, \quad \underline{\varphi}'(s)\underline{\psi}'(s) = \underline{h}_2$$

and that this exhausts all r possible partitions of $\underline{h}_1$ into the disjoint pairs of u-dimensional subvectors and also all r isomorphic partitions of $\underline{h}_2$. Hence

$$\forall q \forall s \forall t: \ \underline{\psi}(s) \cap \underline{\varphi}(t) \neq \Lambda \quad \text{if } s \neq t, \quad \underline{\varphi}'(q) \cap \underline{\psi}'(t) \neq \Lambda \quad \text{if } q \neq t.$$

It follows that the dimensions of the vector $\underline{\psi}(s) \cup \underline{\varphi}(t)$ (respectively $\underline{\varphi}'(q) \cup \underline{\psi}'(t)$) is at most $2u - 1$ and such a vector is a *proper subvector* of the $(2u)$-dimensional vector $\underline{h}_1 = \underline{\varphi}(q)\underline{\psi}(q)$ unless $s = t$ (respectively of the $\underline{h}_2 = \underline{\varphi}'(s)\underline{\psi}'(s)$ unless $q = t$). This proves (6.4). ∎

Now we present our Second Construction. Let m be divided by 9,

$$m = 9v \tag{6.8}$$

and let $\underline{h}_1$, $\underline{h}_2$, $\underline{h}_3$ be the three $(3v)$-dimensional subvectors of $\underline{h}$ that consist of the first $3v$, the next $3v$, and the last $3v$ entries of the vector $\underline{h}$ respectively. Then write that

$$r = (3v)!/(v!)^3 . \tag{6.9}$$

Consider all r possible partitions of the vector $\underline{h}_1$ into the triplets of disjoint v-dimensional subvectors $\underline{\alpha}(s)$, $\underline{\beta}(s)$, $\underline{\gamma}(s)$, $s = 1, \ldots, r$. Let $\underline{\alpha}'(s)$, $\underline{\beta}'(s)$, $\underline{\gamma}'(s)$ and $\underline{\alpha}''(s)$, $\underline{\beta}''(s)$, $\underline{\gamma}''(s)$ be partitions of $\underline{h}_2$ and $\underline{h}_3$ respectively that are isomorphic to the partition $\underline{\alpha}_s$, $\underline{\beta}_s$, $\underline{\gamma}_s$ of $\underline{h}_1$, $s = 1, \ldots, r$.

Then define $\underline{i}(s)$, $\underline{j}(s)$, $\underline{q}(s)$, the vectors-subscripts of Table 5.1 as follows.

$$\underline{i}(s) = \underline{\alpha}(s)\,\underline{\alpha}'(s)\,\underline{\beta}''(s), \quad \underline{j}(s) = \underline{\beta}(s)\,\underline{\gamma}'(s)\,\underline{\alpha}''(s), \quad \underline{k}(s) = \underline{\gamma}(s)\,\underline{\beta}'(s)\,\underline{\gamma}''(s), \quad s = 1, \ldots, r. \tag{6.10}$$

This is our *Second Construction of Generating Tables* 5.1 or the *r-Procedure of TA for* $r = (3v)!/(v!)^3$. Substitute (6.10) in (5.9). Then we obtain

$$\pi(q, s, t) = x^{(q)}_{\underline{\alpha}(q)\underline{\alpha}'(q)\underline{\beta}''(q),\underline{\beta}(q)\underline{\gamma}'(q)\underline{\alpha}''(q)}\, y^{(s)}_{\underline{\beta}(s)\underline{\gamma}'(s)\underline{\alpha}''(s),\underline{\gamma}(s)\underline{\beta}'(s)\underline{\gamma}''(s)}\, z^{(t)}_{\underline{\gamma}(t)\underline{\beta}'(t)\underline{\gamma}''(t),\underline{\alpha}(t)\underline{\alpha}'(t)\underline{\beta}''(t)} . \tag{6.11}$$

Here $1 \le q, s, t \le r$. Equations (6.11) will help us to follow the proof of the next result.

Lemma 6.2. *Let Table 5.1 be defined by the r-Procedure of TA for* $r = (3v)!/(v!)^3$; *see (6.8)–(6.10). Then each correction term of dimension m of that table has degree 1.*

Proof. Let $\pi(q, s, t)$, a correction term of Table 5.1, have dimension m and not have degree 1. Then the v-dimensional vectors $\underline{\gamma}(q)$, $\underline{\alpha}(s)$, $\underline{\beta}(t)$ are to be disjoint. Indeed, if $h(g) \in \underline{\gamma}(q) \cap \underline{\alpha}(s)$ then $h(g) \notin \underline{\alpha}(q) \cup \underline{\beta}(q) \cup \underline{\beta}(s) \cup \underline{\gamma}(s) \cup \underline{h}_2 \cup \underline{h}_3$. Hence the degrees of $x^{(q)}_{\underline{i}(q)\underline{j}(q)}$ and of $y^{(s)}_{\underline{j}(s)\underline{k}(s)}$ in $h(g)$ are equal to 0 (see (6.11) and recall that $\underline{\alpha}(b)$, $\underline{\beta}(b)$, $\underline{\gamma}(b)$ are disjoint for all b, in particular, for $b = q$, $b = s$). If the degree of $z^{(t)}_{\underline{k}(t)\underline{i}(t)}$ in the $h(g)$ is zero then the dimension of $\pi(q, s, t)$ is at most $m - 1$, otherwise the degree of $\pi(q, s, t)$ in the $h(g)$ is one. Hence $\underline{\gamma}(q)$ and $\underline{\alpha}(s)$ are disjoint. Similarly we verify that $\underline{\alpha}(s) \cap \underline{\beta}(t) = \underline{\beta}(t) \cap \underline{\gamma}(q) = \Lambda$. Hence

$$\underline{\gamma}(q)\,\underline{\alpha}(s)\,\underline{\beta}(t) = \underline{h}_1 . \tag{6.12}$$

Similarly we obtain

$$\underline{\beta}'(q)\,\underline{\alpha}'(s)\,\underline{\gamma}'(t) = \underline{h}_2 , \quad \underline{\gamma}''(q)\,\underline{\beta}''(s)\,\underline{\alpha}''(t) = \underline{h}_3 . \tag{6.13}$$

Since the partitions $\underline{\alpha}(\sigma)$, $\underline{\beta}(\sigma)$, $\underline{\gamma}(\sigma)$ of $\underline{h}_1$, $\underline{\alpha}'(\sigma)$, $\underline{\beta}'(\sigma)$, $\underline{\gamma}'(\sigma)$ of $\underline{h}_2$ and $\underline{\alpha}''(\sigma)$, $\underline{\beta}''(\sigma)$, $\underline{\gamma}''(\sigma)$ of $\underline{h}_3$ are isomorphic, (6.13) implies that

$$\underline{\beta}(q)\,\underline{\alpha}(s)\,\underline{\gamma}(t) = \underline{h}_1 , \tag{6.14}$$

$$\underline{\gamma}(q)\,\underline{\beta}(s)\,\underline{\alpha}(t) = \underline{h}_1 . \tag{6.15}$$

Combining (6.12) and (6.14) implies that

$$\underline{\gamma}(q)\,\underline{\beta}(t) = \underline{\beta}(q)\,\underline{\gamma}(t). \tag{6.16}$$

Since for all σ the vectors $\underline{\beta}(\sigma)$, $\underline{\gamma}(\sigma)$ are disjoint and have dimension v, (6.16) implies that

$$\underline{\beta}(q) = \underline{\beta}(t), \quad \underline{\gamma}(q) = \underline{\gamma}(t). \tag{6.17}$$

Similarly (6.12) and (6.15) imply that $\underline{\alpha}(s)\underline{\beta}(t) = \underline{\beta}(s)\underline{\alpha}(t)$ and hence

$$\underline{\alpha}(s) = \underline{\alpha}(t), \quad \underline{\beta}(s) = \underline{\beta}(t). \tag{6.18}$$

Since $\underline{\alpha}(\sigma)\underline{\beta}(\sigma)\underline{\gamma}(\sigma) = \underline{h}_1$ for all σ, we obtain from (6.17), (6.18)

$$\underline{\alpha}(q) = \underline{\alpha}(s) = \underline{\alpha}(t), \quad \underline{\beta}(q) = \underline{\beta}(s) = \underline{\beta}(t), \quad \underline{\gamma}(q) = \underline{\gamma}(s) = \underline{\gamma}(t). \tag{6.19}$$

As follows from the isomorphism of our partitions of $\underline{h}_1$, $\underline{h}_2$, $\underline{h}_3$ and from (6.19), $\pi(q,s,t) = \pi(s,s,s)$ is a *principal term* of Table 5.1. This contradicts our assumption that $\pi(q,s,t)$ is a correction term. ∎

7. Implicit Canceling of Correction Terms of Degree 1 and Resulting Algorithms.

In this section we show how to cancel the correction terms of degree 1 of Table 5.1 defined in the two Constructions of the previous section.

At first, we consider the following class of linear transformations of bilinear problems and algorithms.

Definition 7.1. Let

$$T(\underline{X},\underline{Y},\underline{Z}) = \sum_{\eta} B_\eta(\underline{X},\underline{Y})z_\eta, \quad T^*(\underline{X}^*,\underline{Y}^*,\underline{Z}^*) = \sum_{\eta^*} B_{\eta^*}(\underline{X}^*,\underline{Y}^*)z_{\eta^*}, \tag{7.1}$$

two trilinear forms in $\underline{X}$, $\underline{Y}$, $\underline{Z}$ and in $\underline{X}^*$, $\underline{Y}^*$, $\underline{Z}^*$ respectively define two bilinear problems,

$$\mathcal{B} = \{B_\eta(\underline{X},\underline{Y})\}, \quad \mathcal{B}^* = \{B^*_{\eta^*}(\underline{X}^*,\underline{Y}^*)\} \tag{7.2}$$

respectively. Let a linear transformation

$$\underline{X} = \underline{X}(\underline{X}^*), \quad \underline{Y} = \underline{Y}(\underline{Y}^*), \quad \underline{Z} = \underline{Z}(\underline{Z}^*) \tag{7.3}$$

transform T into T^*, that is

$$T(\underline{X}(\underline{X}^*),\underline{Y}(\underline{Y}^*),\underline{Z}(\underline{Z}^*)) = T^*(\underline{X}^*,\underline{Y}^*,\underline{Z}^*) \tag{7.4}$$

identically in $\underline{X}^*$, $\underline{Y}^*$, $\underline{Z}^*$. Then we write

$$\mathcal{B} = \mathcal{B}(\mathcal{B}^*), \quad T = T(T^*) \tag{7.5}$$

and call $\mathcal{B}$ and T *linear images of* $\mathcal{B}^*$ *and* T^* respectively.

The next illustrative result will not be used in this paper.

Lemma 7.1. *Let (7.1)-(7.5) hold so that* $\mathcal{B} = \mathcal{B}(\mathcal{B}^*)$ *is a linear image of* $\mathcal{B}^*$. *Then (see Definition 2.3)*

$$\rho(\mathcal{B}) \geq \rho(\mathcal{B}^*). \tag{7.6}$$

Proof. Substitute (7.3) in a bilinear algorithm (2.6) of rank M for the problem $\mathcal{B}$. Then (see (7.4))

$$T(\underline{X}, \underline{Y}, \underline{Z}) = T^*(\underline{X}^*, \underline{Y}^*, \underline{Z}^*) = \sum_{q=1}^{M} L_q(\underline{X}(\underline{X}^*))\, L'(\underline{Y}(\underline{Y}^*))\, L''(\underline{Z}(\underline{Z}^*)).$$

This (constructively) defines a bilinear algorithm of rank M for $\mathcal{B}^*$. Choose $M = \rho(\mathcal{B})$ to obtain (7.6). ∎

It is tempting to apply Lemma 7.1 if one seeks upper estimates for $\rho(\mathcal{B}^*)$. Then it would suffice to choose a bilinear problem $\mathcal{B}$ of small rank such that $\mathcal{B}$ is a linear image of $\mathcal{B}^*$. However in the general case we do not have a regular way for the solution of the latter problem. (To appreciate its difficulty, try, for instance, to find a linear transformation which would show that $\mathcal{B} = \mathcal{B}(\mathcal{B}^*)$ in the case $\mathcal{B}^* = \langle m, m, m \rangle$, $\mathcal{B}$ is the PM problem defined by (2.4) where $p = q = m^2$, $\rho(\mathcal{B}) = p + q - 1 = 2m^2 - 1$. If, contrary to our intuition, such a transformation existed then (7.6) would imply that $\rho(\langle m, m, m \rangle) = 2m^2 - 1$, see (2.10).)

Thus we prefer not to use Lemma 7.1. Instead, we will seek for linear transformations that *reduce* the rank of the original algorithms generated by Table 5.1 by canceling the correction terms of degree 1. We call such transformations by *Implicit Canceling (IC)* and the whole process that consists of the choice of Tables 5.1 and of IC by *Trilinear Aggregating with Implicit Canceling (TAIC)*; see [13].

Transformation (7.3) can be considered a triplet of transformations applied to $\underline{X}$, $\underline{Y}$, $\underline{Z}$ separately of each other. In the sequel we apply the transformation (7.3) to the problems $\mathcal{B} = \sum_{s=1}^{r} \oplus \langle I(s), J(s), K(s) \rangle$. In such cases we compose (7.3) of r triplets of linear transformations of $\underline{X}^{*(s)}$, $\underline{Y}^{*(s)}$, $\underline{Z}^{*(s)}$ into $\underline{X}^{(s)}$, $\underline{Y}^{(s)}$, $\underline{Z}^{(s)}$ for all s, $s = 1, \ldots, r$. To simplify the notation, we delete the superscripts s and consider transformations of the triplets $(\underline{X}^*, \underline{Y}^*, \underline{Z}^*)$ into $(\underline{X}, \underline{Y}, \underline{Z})$ and of the trilinear form

$$T(\underline{X}, \underline{Y}, \underline{Z}) = \mathrm{Tr}(XYZ) = \sum_{(\underline{i},\underline{j},\underline{k}) \in D} x_{\underline{i}\underline{j}}\, y_{\underline{j}\underline{k}}\, z_{\underline{k}\underline{i}} \tag{7.7}$$

into another one,

$$T(\underline{X}^*, \underline{Y}^*, \underline{Z}^*) = \mathrm{Tr}(X^*Y^*Z^*) = \sum_{(\underline{i},\underline{j},\underline{k}) \in D^*} x^*_{\underline{i}\underline{j}}\, y^*_{\underline{j}\underline{k}}\, z^*_{\underline{k}\underline{i}}. \tag{7.8}$$

(Recall Remark 5.2.)

Here $\underline{i} = (i(1), \ldots, i(\ell))$, $\underline{j} = (j(1), \ldots, j(\ell'))$, $\underline{k} = (k(1), \ldots, k(\ell''))$ (compare (5.2) (5.4)). The relation $(\underline{i}, \underline{j}, \underline{k}) \in D$ (respectively $(\underline{i}, \underline{j}, \underline{k}) \in D^*$) under the sign $\sum$ designates the summation in

$i(1),\dots,i(\ell)$, $j(1),\dots,j(\ell')$, $k(1),\dots,k(\ell'')$ from 0 (respectively 1) to $n-1$. The latter comments also define two domains, D and D^* where the $\underline{i}$, $\underline{j}$, $\underline{k}$ range.

The trilinear forms of (7.7), (7.8) define the problems $\langle I,J,K\rangle$ and $\langle I^*,J^*,K^*\rangle$ respectively where

$$I=n^{\ell},\quad J=n^{\ell'},\quad K=n^{\ell''},\quad H=IJK=n^{m}. \tag{7.9}$$

$$I^*=(n-1)^{\ell},\quad J^*=(n-1)^{\ell'},\quad K^*=(n-1)^{\ell''},\quad H^*=I^*J^*K^*=(n-1)^{m}. \tag{7.10}$$

Here is one of possible linear transformations of (7.7) into (7.8).

$$x_{\underline{i}\underline{j}}=x^*_{\underline{i}\underline{j}},\quad y_{\underline{j}\underline{k}}=y^*_{\underline{j}\underline{k}},\quad z_{\underline{k}\underline{i}}=z^*_{\underline{k}\underline{i}}\quad \text{for } (\underline{i},\underline{j},\underline{k})\in D^*, \tag{7.11}$$

$$z_{\underline{k}\underline{i}}=0\quad \text{if } k(t'')=0,\qquad \sum_{k(t'')=0}^{n-1} y_{\underline{j}\underline{k}}=0, \tag{7.12}$$

$$y_{\underline{j}\underline{k}}=0\quad \text{if } j(t')=0,\qquad \sum_{j(t')=0}^{n-1} x_{\underline{i}\underline{j}}=0, \tag{7.13}$$

$$z_{\underline{k}\underline{i}}=0\quad \text{if } i(t)=0,\qquad \sum_{i(t)=0}^{n-1} x_{\underline{i}\underline{j}}=0, \tag{7.14}$$

We assume that all unbounded entries of $\underline{i}$, $\underline{j}$, $\underline{k}$ that are used in (7.12)-(7.14) range in the domain D and that t, t', t'' range as follows, $t''=1,\dots,\ell''$ in (7.12), $t'=1,\dots,\ell'$ in (7.13), and $t=1,\dots,\ell$ in (7.14).

Equations (7.11)-(7.14) contain some implicit expressions of $x_{\underline{i}\underline{j}}$ and $y_{\underline{j}\underline{k}}$ as linear functions of $\underline{X}^*$, $\underline{Y}^*$. To make them explicit, rewrite the second equations of (7.12) (7.14) so that for each triplet t, t', t'' all indeterminates are moved to the right sides except the following ones which remain in the left sides,

$y_{\underline{j}\underline{k}}$ where $k(t'')=0$ in (7.12),
$x_{\underline{i}\underline{j}}$ where $j(t')=0$ in (7.13),
$x_{\underline{i}\underline{j}}$ where $i(t)=0$ in (7.14).

Then substitute (7.11) in the right sides.

Now apply a variation of the linear transformation (7.11) (7.14) to each of the r triplets $\underline{X}=\underline{X}^{(s)}$, $\underline{Y}=\underline{Y}^{(s)}$, $\underline{Z}=\underline{Z}^{(s)}$, $s=1,\dots,r$ of indeterminates of Table 5.1 defined by our First Construction of Section 6. In that variation preserve (7.11) (7.14) for all t, t'' and also for all $t'\le u$ (then $j(t')\in \underline{h}_1$). If $t'>u$ (then $j(t')\in \underline{h}_2$) substitute the following equations for (7.13),

$$x_{\underline{i}\underline{j}}=0 \text{ if } j(t')=0,\qquad \sum_{j(t')=0}^{n-1} y_{\underline{j}\underline{k}}=0,\quad t'>u. \tag{7.15}$$

Notice that, by virtue of Lemmas 5.2, 6.1, the above transformation applied to the First Construction of Section 6 cancels all correction terms of Table 5.1. This gives us the following estimate; see (5.10), (6.2), (7.9), (7.10).

Theorem 7.1. *For arbitrary natural numbers u and n,*

$$[(2u)!/(u!)^2] \odot \langle (n-1)^u, (n-1)^{2u}, (n-1)^u) \leftarrow : n^{4u} \odot \langle 1, 1, 1 \rangle \tag{7.16}$$

We will call the transformation (7.11)-(7.15) the *First Transformation for Implicit Canceling.* The associated equation of (7.16) for a fixed n and sufficiently large u implies the following estimate (see Theorem 3.1).

Corollary 7.1. *For arbitrary natural n, $\beta(n) = 3(2\log n - \log 2)/2\log(n-1)$ is a limiting exponent of MM, in particular, $\beta(9) < 2.67$ is a limiting exponent of MM.*

Next we define our second linear transformation which also transforms (7.7) into (7.8) and enables us to cancel all correction terms of degree 1 in any Table 5.1.

We define this transformation recursively in m steps. With each step we associate a new value of t'', t' or t. For instance, we can successively choose $t'' = 1, \ldots, \ell''$, then $t' = 1, \ldots, \ell'$, then $t = 1, \ldots, \ell$.

Here is the first step of the transformation in the case $\ell'' = 1$ where we designate $k = k(1) = \underline{k}$.

$$\forall \underline{i} \forall \underline{j}: \; x_{\underline{i}\underline{j}} = x^*_{\underline{i}\underline{j}}. \tag{7.17}$$

$$\forall \underline{j} \forall k \; (k \neq 0): \; y_{\underline{j}k} = y^*_{\underline{j}k}. \tag{7.18}$$

$$\forall \underline{i} \forall \underline{j}: \; \sum_{h=0}^{n-1} y_{\underline{j}h} = \sum_{h=0}^{n-1} z_{h\underline{i}} = 0. \tag{7.19}$$

$$\forall \underline{i} \forall k \; (k \neq 0): \; z_{k\underline{i}} + \sum_{h=1}^{n-1} z_{h\underline{i}} = z^*_{k\underline{i}}. \tag{7.20}$$

Equations (7.19) contain implicit expressions of $y_{\underline{j}0}$, $z_{0\underline{i}}$ through $\{y^*_{\underline{j}k}, z^*_{k\underline{i}}, k = 1, \ldots, n-1\}$ which can be easily turned into explicit ones. Similarly Equations (7.20) implicitly express $z_{k\underline{i}}$ as linear function of $z^*_{k\underline{i}}$ for $k = 1, \ldots, n-1$. To obtain the explicit expressions, we have to solve (7.20) over F for each $\underline{i}$ as a system of linear equations in $z_{k\underline{i}}$, $k = 1, \ldots, n-1$. The next simple result shows that the solution exists if $n \neq 0$ in F.

Lemma 7.2. *For each $\underline{i}$ the determinant of the system of Equations (7.20) in $z_{1\underline{i}}, \ldots, z_{n-1,\underline{i}}$ is equal to n.*

Next we examine how the transformation (7.17)-(7.20) change the trilinear form $T(\underline{X}, \underline{Y}, \underline{Z})$. We write that

$$T = T(\underline{X}, \underline{Y}, \underline{Z}) = \sum_{\underline{i},\underline{j}} \sum_{k=0}^{n-1} x_{\underline{i}\underline{j}} \, y_{\underline{j}k} \, z_{k\underline{i}}$$

$$= \sum_{\underline{i},\underline{j}} x_{\underline{i}\underline{j}} \Big(\sum_{k=1}^{n-1} y_{\underline{j}k} \, z_{k\underline{i}} + y_{\underline{j}0} \, z_{0\underline{i}} \Big).$$

Substitute $y_{\underline{j}0} = -\sum_{k=1}^{n-1} y_{\underline{j}k}$ (see (7.19)) and obtain

$$T = \sum_{\underline{i},\underline{j}} x_{\underline{i}\underline{j}} \sum_{k=1}^{n-1} y_{\underline{j}k}(z_{k\underline{i}} - z_{0\underline{i}}).$$

Then substitute $z_{0\underline{i}} = -\sum_{h=1}^{n-1} z_{h\underline{i}}$ (see (7.19)). This gives

$$T = \sum_{\underline{i},\underline{j}} x_{\underline{i}\underline{j}} \sum_{k=1}^{n-1} y_{\underline{j}k}\Big(z_{k\underline{i}} + \sum_{h=1}^{n-1} z_{h\underline{i}}\Big). \tag{7.21}$$

Substitute (7.17), (7.18), (7.20) in (7.21) and obtain

$$T = T(\underline{X},\underline{Y},\underline{Z}) = \sum_{\underline{i},\underline{j}} \sum_{k=1}^{n-1} x^*_{\underline{i}\underline{j}}\, y^*_{\underline{j}k}\, z^*_{k\underline{i}} = T(\underline{X}^*,\underline{Y}^*,\underline{Z}^*).$$

We come to the following result.

Lemma 7.3. *For arbitrary ℓ, ℓ', n ($n \neq 0$ in F) the linear transformation (7.17)–(7.20) transforms $\langle n^\ell, n^{\ell'}, n\rangle$ into $\langle n^\ell, n^{\ell'}, n-1\rangle$.*

In the case $\ell'' > 1$ we can generalize (7.17)–(7.20) using the following notation.

Notation 7.1. Delete the entry $k(t'')$ of the vector $\underline{k}$. Designate the resulting vector by $\underline{k}(t'')$. Designate $\underline{k} = \underline{k}(t'')k(t'')$ in the case where all entries of $\underline{k}$ are considered integer parameters. If the value of $k(t'')$ is fixed, $k(t'') = h$ and if other entries of $\underline{k}$ are parameters, designate $\underline{k} = \underline{k}(t'')h$.

Then the transformation (7.17)–(7.20) can be generalized to the case $\ell'' > 1$ where t'' is fixed, $1 \leq t'' \leq \ell''$. Let (7.17) be preserved and the following equations substitute for (7.18)–(7.20).

$$\forall \underline{j}\, \forall \underline{k}(t'')\, \forall k(t'')\, (k(t'') \neq 0)\colon\ y_{\underline{j}\underline{k}} = y^*_{\underline{j}\underline{k}}. \tag{7.22}$$

$$\forall \underline{i}\, \forall \underline{j}\, \forall \underline{k}(t'')\colon\ \sum_{h=0}^{n-1} y_{\underline{j},\underline{k}(t'')h} = \sum_{h=0}^{n-1} z_{\underline{k}(t'')h,\underline{i}} = 0. \tag{7.23}$$

$$\forall \underline{i}\, \forall \underline{j}\, \forall \underline{k}(t'')\ (k(t'') \neq 0)\colon\ z_{k\underline{i}} + \sum_{h=1}^{n-1} z_{\underline{k}(t'')h,\underline{i}} = z^*_{k\underline{i}}. \tag{7.24}$$

Remark 7.1. If $\sum_{h=0}^{n-1} z^*_{\underline{k}(q'')h,\underline{i}} = 0$ and (7.24) holds then $\sum_{h=0}^{n-1} z_{\underline{k}(q'')h,\underline{i}} = 0$ for any q'', $1 \leq q'' \leq \ell''$, $q'' \neq t''$.

Then similarly to Lemma 7.3 the following result can be obtained.

Lemma 7.4. *For arbitrary ℓ, ℓ', ℓ'', n ($n \neq 0$ in F) the linear transformation (7.17), (7.22)–(7.24) transforms $\langle n^\ell, n^{\ell'}, n^{\ell''}\rangle$ into $\langle n^\ell, n^{\ell'}, n^{\ell''-1}(n-1)\rangle$. Similarly $\langle n^\ell, n^{\ell'}, n^{\ell''}\rangle$ can be transformed into $\langle n^\ell, n^{\ell'-1}(n-1), n^{\ell''}\rangle$ and into $\langle n^{\ell-1}(n-1), n^{\ell'}, n^{\ell''}\rangle$.*

Recursively applying the three latter transformations we obtain the desired linear functions (7.3) that for arbitrary $n \neq 0$, ℓ, ℓ', ℓ'' transform $\langle n^{\ell}, n^{\ell'}, n^{\ell''} \rangle$ into $\langle (n-1)^{\ell}, (n-1)^{\ell'}, (n-1)^{\ell''} \rangle$. We call such a process the *Second Transformation for Implicit Canceling*. Its efficiency stems from the following fact which can be easily verified using Remark 7.1 and similar observations.

Lemma 7.5. *Let functions (7.3) define the Second Transformation for Implicit Canceling. Then (7.23) holds for all t'', $t'' = 1, \ldots, \ell''$ as well as the following equations.*

$$\forall \underline{j} \, \forall \underline{k} \, \forall \underline{i}(t): \quad \sum_{h=0}^{n-1} x_{\underline{i}(t)h, \underline{j}} = \sum_{h=0}^{n-1} z_{\underline{k}, \underline{i}(t)h} = 0, \quad t = 1, \ldots, \ell, \tag{7.25}$$

$$\forall \underline{k} \, \forall \underline{i} \, \forall \underline{j}(t'): \quad \sum_{h=0}^{n-1} x_{\underline{i}, \underline{j}(t')h} = \sum_{h=0}^{n-1} y_{\underline{j}(t')h, \underline{k}} = 0, \quad t' = 1, \ldots, \ell', \tag{7.26}$$

Corollary 7.2. *Let the Second Transformation for IC be applied to an arbitrary Table 5.1 then Equations (7.23), (7.25), (7.26) cancel all correction terms of degree 1.*

(Corollary 7.2 follows from Lemmas 5.2, 7.5.)

In particular, if Table 5.1 is defined by the First Construction of Section 6 then all correction terms of Table 5.1 are canceled. This gives another proof of (7.16) (for $n \neq 0$ in F). If Table 5.1 is defined by the Second Construction of Section 6 then only the correction terms of dimensions at most $m-1$ are not canceled by the Second Transformation for IC. This gives the following result.

Corollary 7.3. *For arbitrary field F and natural v, n ($n \neq 0$ in F)*

$$\frac{(3v)!}{(v!)^3} \odot \langle (n-1)^{3v}, (n-1)^{3v}, (n-1)^{3v} \rangle \leftarrow: (n^{9v} + \rho c^*) \odot \langle 1, 1, 1 \rangle .$$

where ρc^* is the rank of the sum of all instances of all correction terms of Table 5.1 transformed by the Second Transformation for IC. Here Table 5.1 is defined by the Second Construction of Section 6.

Our next objective is the following estimate.

Lemma 7.6. *Under the conditions of Corollary 7.3,*

$$n^{9v} + \rho c^* \leq (n+1)^{9v}. \tag{7.27}$$

Proof. Let Table 5.1 be defined by the Second Construction of Section 6. Let the Second Transformation for IC be applied. Then for all μ consider all possible sets of μ different integers

$$G = \{g_1, \ldots, g_\mu\}, \quad 1 \leq g_\eta \leq 9v, \quad \eta = 1, \ldots, \mu, \quad \mu = 0, 1, \ldots, 9v. \tag{7.28}$$

. Let one of such sets be fixed. Substitute zeroes for each indeterminate $x^{(s)}_{\underline{i}(s)\underline{j}(s)}$, $y^{(s)}_{\underline{j}(s)\underline{k}(s)}$, $z^{(s)}_{\underline{k}(s)\underline{i}(s)}$, in Table 5.1 unless such an indeterminate has degree zero in $h(g_\eta)$ for $\eta = 1, \ldots, \mu$. Call the

resulting table by the *Auxiliary Table associated with the set* $\{g_1,\dots,g_\mu\}$. (Table 5.1 itself is associated with the empty set.) Notice that for $\mu \geq 1$ all principal terms of all Auxiliary Tables are zeroes.

Multiply the aggregate of the Auxiliary Table associated with the set $\{g_1,\dots,g_\mu\}$ by $(-n)^\mu$. Sum the results for all values of all entries $h(g) \in \underline{h}$ such that $g \notin \{g_1,\dots,g_\mu\}$ and for all possible sets $\{g_1,\dots,g_\mu\}$, $\mu = 0,1,\dots,9v$. As can be verified, no correction terms of dimensions less than m remain in the resulting total sum. Hence the sum is identically $T(\underline{X},\underline{Y},\underline{Z})$ because the correction terms of dimension m are canceled, by virtue of Lemmas 5.2, 6.2, 7.5. It remains to estimate $n^{9v} + \rho c^*$, the rank of the sum of all instances of all aggregates in all of our Auxiliary Tables in order to prove (7.27). (This whole procedure for canceling the terms of dimensions less than m is general. It can be called the *Alternating Summation of Aggregates.*)

The desired upper estimate (7.27) can be obtained from the next two simple lemmas.

Lemma 7.7. *For a natural* μ, $0 \leq \mu \leq 9v$, *and for an Auxiliary Table associated with a set* $\{g_1,\dots,g_\mu\}$ *(see (7.28)) there exist at most* $n^{9v-\mu}$ *instances of the aggregate of that table.*

Lemma 7.8. *For an arbitrary natural* μ, $0 \leq \mu \leq 9v$, *there exist at most* $\binom{9v}{\mu} = (9v)/(\mu!(9v-\mu)!)$ *different sets* $\{g_1,\dots,g_\mu\}$ *where* g_η *are natural numbers,* $1 \leq g_\eta \leq 9v$.

Corollary 7.4. *For arbitrary field of constants* F *and for all natural* v, n, $(n \neq 0$ *in* $F)$, *the following mapping is valid.*

$$\frac{(3v)!}{(v!)^3} \odot ((n-1)^{3v},(n-1)^{3v},(n-1)^{3v}) \leftarrow (n+1)^{9v} \odot (1,1,1)\,.$$

The associated equations for a fixed n and for $v \to \infty$ define the limiting exponents of MM,

$$\beta^*(n) = \log((n+1)^3/3)/\log(n-1)\,, \tag{7.29}$$

in particular,

$$\beta^*(20) < 2.7288\,.$$

8. Conclusions.

How can the techniques of the previous sections be improved? One of the natural ways is to improve the Constructions of Section 6.

Corollary 7.2 enables us to cancel all correction terms of degree 1. The method of the Alternating Summation of Aggregates (see the proof of Lemma 7.6) can be generalized for canceling the terms of dimensions less than m. It remains to design Generating Table 5.1 where all correction terms of dimension m would have degree 1 in some of the $h(g)$. We proved such a property for the Second Construction of Section 6. The proof and hence the result itself can be extended to any Table 5.1 such that the vectors of subscripts $\underline{k}(q)$, $\underline{i}(s)$, $\underline{j}(t)$ are disjoint only if $q = s = t$.

Is it possible to obtain Table 5.1 with r rows where the latter property holds and where $3(m \log n - \log r)/m \log(n-1)$ is substantially less than $\beta^*(n)$ in (7.29)? (See (5.1), (7.9), (7.10), (7.29).)

Here is another way that seems to be more promising. One can generalize Tables 5.1 by turning them into the following ones which we call *Generating* λ*-Tables*. (We represent only the s-th row of the tables, assuming that $s = 1, \ldots, r$.)

Table 8.1.

$$\alpha(s,\lambda)x^{(s)}_{\underline{i}(s)\underline{j}(s)} \qquad \beta(s,\lambda)y^{(s)}_{\underline{j}(s)\underline{k}(s)} \qquad \gamma(s,\lambda)z^{(s)}_{\underline{k}(s)\underline{i}(s)}$$

Here $\alpha(s,\lambda)$, $\beta(s,\lambda)$, $\gamma(s,\lambda)$ are constants of F such that

$$\forall s\colon \sum_{\lambda} \alpha(s,\lambda)\,\beta(s,\lambda)\,\gamma(s,\lambda) = 1\,.$$

We assume that the aggregates of Table 8.1 are to be summed for all values of λ. (In particular, if λ is a constant and $\alpha(s,\lambda) = \beta(s,\lambda) = \gamma(s,\lambda) = 1$ for all s, then we come back to Table 5.1.) The coefficients $\alpha(s,\lambda)$, $\beta(s,\lambda)$, $\gamma(s,\lambda)$ can be chosen such that several correction terms are canceled in the result of the summation in λ. More precisely, it is sufficient to satisfy the equation

$$\sum_{\lambda} \alpha(q,\lambda)\,\beta(s,\lambda)\,\gamma(t,\lambda) = 0 \tag{8.1}$$

in order to cancel the correction term $\pi_\lambda(q,s,t)$,

$$\sum_{\lambda} \pi_\lambda(q,s,t) = 0\,. \tag{8.2}$$

In particular, in some cases this observation enables us to cancel even the correction terms whose degrees in all $h(g)$ are greater than 1 (if such terms appear in Table 8.1).

In fact, such a trick was successfully applied in [3, 12] under the name Trilinear Canceling (see also [9]). On the other hand, the Generating λ-Tables can be used to define λ-algorithms for MM which turn out to coincide with APA-algorithms if $\alpha(s,\lambda)$, $\beta(s,\lambda)$, $\gamma(s,\lambda)$ are rational functions of λ and if the consideration is modulo λ. In such a setting the application of (8.1), (8.2) as a means of canceling is generally efficient. This is formally proven in the basic theorem on the relations between usual algorithms and APA-algorithms. (Such an interpretation of the theorem can be seen from the original illuminating proof given in [6] and repeated in neither of the papers [7 10, 17].) During the study of APA-algorithms this direction has remained in the shadows. However regarding the relationship between APA-algorithms and λ-Tables the approach of [6] seems important and might become fruitful in the future.

In particular, it is important to understand the most efficient ways of canceling the correction terms of Generating λ-Tables. It might happen that the existent methods already rely on nearly optimum ways of such a canceling. However because of the extreme irregularity of the asymptotically fastest known algorithms for MM we might be far from understanding the successful methods of canceling hidden in those algorithms. Then further efforts in the analysis of the best existent methods of MM can become fruitful.

Acknowledgments.

I wish to thank Prof. D. E. Knuth for his stimulating interest to the MM problem.

References

[1] V. Strassen, "Gaussian elimination is not optimal," *Numer. Math.* **13**, (1969), 354–356.

[2] V. Ya. Pan, (unpublished manuscript) (1965).

[3] V. Ya. Pan, "Strassen's algorithm is not optimal," *Proc. 19th Annual Symposium on Foundations of Computer Science*, (1978), 166–176.

[4] V. Ya. Pan, "On schemes for the computation of products and inverses of matrices," (in Russian), *Russian Math. Survey* **27** 5 (1972), 249–250.

[5] D. Bini, M. Capovani, G. Lotti, F. Romani, "$O(N^{2.7799})$ complexity for matrix multiplication," *Information Processing Letters* **8** 5 (1979), 234–235.

[6] D. Bini, "Relations between EC-algorithms and APA-algorithms," *Applications*, Calcolo XVII (1980), 87–97.

[7] A. Schönhage, "Partial and total matrix multiplication," TR, University of Tübingen, (1979).

[8] V. Ya. Pan, "Field extension and trilinear aggregating, uniting and canceling for the acceleration of matrix multiplication," *Proc. 20th Annual Symposium on Foundations of Computer Science*, San Juan, Puerto Rico, (1979), 28–38.

[9] V. Ya. Pan, "New combinations of methods for the acceleration of matrix multiplication," *Computers and Math. (with Appl.)* **7** 1 (1981), 73–125.

[10] A. Schönhage, *SIAM J. on Computing* **10** 3, (1981), 434–456.

[11] V. Strassen, "Vermeidung von division," *J. Reine Angew. Math.* **264** (1973), 184–202.

[12] V. Ya. Pan, "New fast algorithms for matrix operations," *SIAM J. on Computing* **9** 2 (1980), 321–342.

[13] V. Ya. Pan, "Trilinear aggregating with implicit canceling for a new acceleration of matrix multiplication," *Computers and Math. (with Appl.)*, in press.

[14] V. Ya. Pan and S. Winograd, "The book of abstracts of Oberwolfach Conferences," Oberwolfach (October 26, 1979).

[15] *Bulletin of the EATCS* **10** (1980), 99–100.

[16] F. Romani, "Some properties of disjoint sums of tensors related to matrix multiplication," *Nota Interna* B 80/4, *IEI, Pisa*, Italy (1980).

[17] D. Coppersmith and S. Winograd, "On the asymptotic complexity of matrix multiplication," IBM T. J. Watson Research Center, (1981).

[18] V. Strassen, "Evaluation of Rational Functions," in *Complexity of Computer Computations*, (R. E. Miller and J. W. Thatcher, Eds.), Plenum Press, New York, (1972), 1–10.

[19] A. V. Aho, J. E. Hopcroft, and J. D. Ullman, *The Design and Analysis of Computer Algorithms*, Addison-Wesley (1976).

[20] A. Borodin and I. Munro, *The Computational Complexity of Algebraic and Numeric Problems*, American Elsevier (1975).

[21] C. M. Fiduccia and Y. Zalstein, "Algebras having linear multiplicative complexity," *J. ACM* **24** 2 (1977), 311–331.

[22] D. E. Knuth, *The Art of Computer Programming: Seminumerical Algorithms* **2**, Addison-Wesley (1981).

[23] S. Winograd, *Arithmetic complexity of computations*, SIAM (1980).

[24] R. W. Brockett and D. Dobkin, "On the optimal evaluation of a set of bilinear forms," *Linear Algebra and Appl.* **19** (1978), 207–235.

[25] J. C. Lafon and S. Winograd, "A lower bound for the multiplicative complexity of the product of two matrices," *Theoretical Computer Science*, in press.


DATE:
ILME